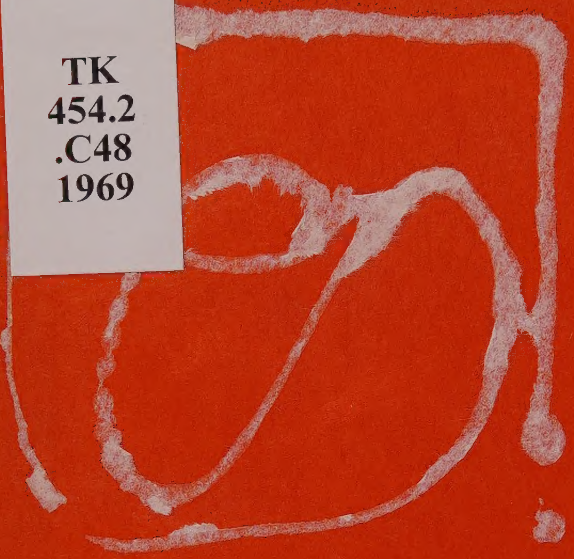
LEON O. CHUA INTRODUCTION TO NONLINEAR NETWORK THEORY

TABOR COLLEGE LIBRARY
[redacted] main
Chua, Leon O./Introduction to nonlinear
3 2002 00034637 7

TK
454.2
.C48
1969



DATE DUE

~~MAY 7 1979~~

INTRODUCTION TO
NONLINEAR NETWORK THEORY

McGRAW-HILL SERIES IN ELECTRONIC SYSTEMS

John G. Truxal and Ronald A. Rohrer, Consulting Editors

CHUA: Introduction to Nonlinear Network Theory
HUELSMAN: Theory and Design of Active RC Circuits
MEDITCH: Stochastic Optimal Linear Estimation and Control
MELSA AND SCHULTZ: Linear Control Systems
ROHRER: Introduction to the State Variable Approach to Network Theory
SCHULTZ AND MELSA: State Functions and Linear Control Systems
STAGG AND EL-ABIAD: Computer Methods in Power System Analysis
TIMOTHY AND BONA: State Space Analysis: An Introduction

621.3
C559i

INTRODUCTION TO
NONLINEAR NETWORK THEORY

LEON O. CHUA
Associate Professor of Electrical Engineering
Purdue University

734651

TABOR COLLEGE LIBRARY
Hillsboro, Kansas 67063

McGRAW-HILL BOOK COMPANY
New York St. Louis San Francisco London
Sydney Toronto Mexico Panama

Introduction to Nonlinear Network Theory

Copyright © 1969 by McGraw-Hill, Inc. All rights reserved.
Printed in the United States of America. No part of this
publication may be reproduced, stored in a retrieval system,
or transmitted, in any form or by any means, electronic,
mechanical, photocopying, recording, or otherwise, without
the prior written permission of the publisher.

Library of Congress Catalog Card Number 76-75165

ISBN 07-010804-

2 3 4 5 6 7 8 9 0 MAMB 7 6 5 4 3 2 1

To my wife Diana
with love and gratitude

PREFACE

In recent years, teachers of electronic circuit theory have been faced with the dilemma of a decreasing amount of time in which to present circuits containing an increasing variety of modern electronic devices. This problem has been brought on by two parallel developments, the tremendous advances in material science and technology and the equally significant advances in computer science and numerical techniques. The first of these forward strides has resulted not only in the invention and development of many new devices, but also in the advent of large-scale integration. The integrated circuit, in which thousands of transistors may be fabricated in a single monolithic chip, has added a new dimension to electronic circuit analysis. On the other hand, progress in computer sciences has greatly influenced the growing trend toward a systems approach to engineering education, concurrently forcing more conventional circuit courses out of the curriculum in the competition for the students limited time. The result of these developments, whether we like it or not, is the end of what some might consider the "era" of conventional electronic circuit theory.

In the effort to teach more in less time, a more constructive approach would seemingly be to attempt a *unification of existing theories.* Such an approach is what this book is all about.

To implement this new approach, the electrical engineering group at Purdue has rearranged conventional topics in circuit theory into two areas: linear active networks and nonlinear networks. The course in linear networks covers small-signal networks for which linear incremental models are valid, including small-signal amplifiers and simple active filters. The nonlinear presentation, of course, covers networks which are not linear. This arrangement was chosen because there is a well-established theory for analyzing linear networks, but no such currently available general theory for analyzing nonlinear networks.

Because of this lack of general theory, the usual approach to nonlinear networks has been to collect, to catalog, and to refer to standard "recipes" for analysis of the various types of networks. However, since the student learns only recipes for analyzing specific circuits containing specific devices, he is not prepared to work with nonlinear circuits or devices he has not seen before (unless he is the fortunate individual who is creative enough to develop his own new recipe). When we consider the rapid development of new devices, it does not take much imagination to see that even an engineer fresh out of the university may find himself in the same serious predicament as the engineers who studied vacuum tubes in school and had to confront the transistor on their own.

To rectify the shortcomings of the conventional approach, in this book we have *emphasized the principles involved in composing the recipes, not the recipes themselves.* Accordingly, *most of the methods developed in this book are device-independent,* a situation made possible by what we call the *black-box approach* to electronic circuit analysis. Essentially, what we are saying is that any electronic device is considered to be a black box with two or more electrical terminals accessible to the external world. The box is black in the sense that we do not know its contents, and indeed, as far as circuit theory is concerned, they do not matter to us. We will never ask why a device has a certain characteristic, for such a question can be answered—if at all—only by looking inside the box to discover its internal structure; in other words, we will not study the physics of the device. In fact, if two black boxes have identical electrical characteristics, we do not care if one contains a familiar solid-state device and the other a piece of rock from the moon. In other words, we will not study the physics of the device (after all, we need not learn semiconductor physics to analyze a transistor circuit). The important study of the physics of devices belongs to a course in material science, not circuit theory.

In line with the presentation of principles and concepts in as general a context as possible, we have found it necessary to take a deductive approach on many occasions; that is, each principle is first discussed in general terms, followed by various special cases and a number of examples. Because the methods presented are very general, they are not necessarily the easiest to apply in any particular situation. However, we believe that *it is much more rewarding for a student to learn at least one general method that al-*

ways works, than to learn a number of simpler (but more specialized) methods that hold in one case but not in another.

We trust that by learning the general principles presented in this book, the student will be able to develop certain shortcuts to specific problems. At the very least, we expect him to have developed enough confidence to handle any new device (after all, it is just another black box) and to analyze simple nonlinear circuits.

It should be noted here that we do not expect our student to solve a problem in the shortest possible time. We expect our student to know the concepts, we do not expect him to be skillful; the skill will be developed later in a design-oriented course.

Although a considerable amount of material in this book is devoted to the synthesis problem, we will study only the basic elements of nonlinear network synthesis and will stop as soon as a basic "skeleton" circuit is derived. The task of optimizing the circuits performance is also left to the design-oriented course.

As far as actual content is concerned, this book is divided into three parts. Part 1 contains four chapters of *foundation material*. The first three chapters are devoted to the definition and characterization of nonlinear network elements. A number of practical applications are presented for motivational purposes. The mature reader may skim over this material or even skip it completely. Chapter 4 develops the general procedures for writing the equations of motion for nonlinear networks. The universe of all networks is divided into the class of *resistive nonlinear networks* and the class of *dynamic nonlinear networks.* Basic numerical techniques are introduced in this chapter for solving these two classes of nonlinear networks on a computer.

Part 2 contains eight chapters devoted exclusively to resistive nonlinear networks. Chapter 5 defines three basic concepts of resistive networks, namely, the operating point, the driving-point plot, and the transfer-characteristic plot. In view of their frequent appearance, we have found it convenient to abbreviate the last two terms by DP plot and TC plot, respectively. Chapters 6 and 8 are the most basic chapters in Part 2; the former deals with analysis of nonlinear networks, whereas the latter deals with the synthesis of nonlinear networks. In order to extend the methods of Chapter 6 to a larger class of networks, a number of nonlinear network theorems and techniques are required; these are presented in Chapter 7. A number of common-functional nonlinear circuits normally taught in higher level courses are presented in Chapters 9 and 10 with a very unconventional approach. These chapters are included to illustrate the methods presented in Chapters 6, 7, and 8. Although no new concepts are introduced, our experience has shown that these two chapters have high motivational value. Chapter 11 presents the general principles for modeling three-terminal nonlinear elements in terms of two-terminal elements and controlled sources. This chapter leads to the formulation of a completely general method for analyzing nonlinear resistive networks, as presented in Chapter 12. Ideally, the digital computer should be used to implement the methods in this chapter.

Part 3, consisting of eight chapters, develops the theory and analysis of dynamic nonlinear networks. Chapter 13 introduces the basic concepts useful in dynamic network analysis. Here the state space is presented from a geometric point of view. Several powerful techniques for analyzing autonomous first-order nonlinear networks are presented in Chapter 14. These techniques are then generalized in Chapter 15 to the case of nonlinear switching networks. The techniques of Chapter 15 are used to present a unifying theory of first-order multivibrators and time-base generators in Chapters 16 and 17, respectively. Several new principles for analyzing first-order nonautonomous networks and second-order autonomous networks are given in Chapters 18 and 19, respectively. Finally, the book concludes with a discussion of *n*th-order nonlinear network analysis. This chapter serves as a simple introduction to a more advanced nonlinear network theory course requiring higher mathematics.

Most chapters start with elementary and basic concepts and are followed by more general theories, algorithms, and their applications. The last few sections of each chapter are usually of a more specialized or advanced nature, and can be omitted on first reading without loss of continuity. In particular, the following sections fall into this category: 1-9, 1-10, 3-3, 3-4, 3-7, 3-8 , 4-5, 4-7, 4-8, 5-4, 6-7, 6-8, 7-7, 7-8, 7-9, 8-2-3, 8-5, 12-6, 13-4-3, 14-7, 14-8-3, 15-4, 19-4, 19-5, 19-6, 20-2, and 20-3.

Exercises are included after most sections. These pertain to the immediately preceding material. Some of the exercises are drill-type questions designed to test or clarify new concepts introduced in a particular section and can usually be answered with little or no calculation. Other exercises may involve the derivation of related results or the working of specific examples designed to illustrate the concepts and algorithms presented in a particular section.

Problems appear at the end of each chapter. These are plentiful and sometimes very long; a selected group of them is sufficient for any course. In order to enhance their value to the book, many of the problems contain several *similar* parts, so that only one part need be assigned at a time. Although all problems could be solved without using a computer, a few of them involve such considerable computation that it would be desirable to use a computer.

A package of computer subroutines and programs has been carefully designed to implement many of the basic algorithms described in the text. These programs and subroutines are written in FORTRAN IV and are currently available from the author. These subroutines and programs are developed strictly from a pedagogical point of view; since they are designed to eliminate only routine calculations and tedious graphical constructions, the student must learn the conceptual aspects of each algorithm before he can successfully use the package. A more detailed description of the basic philosophy and mechanics of this package is given elsewhere.[1]

This book is virtually self-contained; *all prerequisite and background material is carefully developed as needed. No mathematical background beyond that of basic calculus is assumed.* Moreover, the material has been developed so that a linear active network theory course is unnecessary either as a prerequisite or as a corequisite. (*Familiarity* with the basic characteristics and operating principles of electronic devices is helpful but not at all necessary.) In fact, selected portions of this book can be used as a first course in circuit theory. After obtaining a solid foundation in nonlinear circuits, the student can then take a course in linear circuit theory in which powerful analytical techniques can be introduced. Then not only will the student better appreciate these materials, he will also be more alert to the limitations of linear techniques. Another compelling reason for teaching the present text before a linear circuits course is to give the student at least one more semester to take some basic mathematics courses (preferably matrix theory and differential equations) so he can better understand the state-space approach in linear circuits.

With an appropriate selection of subject matter, this book could be used as a textbook in electronic circuit theory at the sophomore, junior, or senior levels. For a three-credit-hour sophomore course, the topics might consist of only a few introductory sections (such as 1-6, 2-1-1, 3-2, 4-1, and 4-2) from Part 1. These might be followed by a heavy dose of selected topics from Chapters 6 through 12 in Part 2. The third part of the course might consist of the more basic topics from Chapters 13 through 18. A three-credit hour junior course might cover essentially the same topics proposed for the sophomore course but in greater depth. This would include additional sections previously omitted. It is reasonable to assign more difficult homework problems at this level. At the senior level, the entire book can be covered by assigning the first five chapters as review material. Since this book contains much new material in the general area of nonlinear circuit analysis and synthesis, it can also be used as a reference book by practicing engineers involved in the design of large-signal electronic circuits.

---

[1] Leon O. Chua, A Computer-Oriented Sophomore Course in Nonlinear Network Theory, *IEEE Trans. Education,* Special Issue on Computer Application in Education, September, 1969.

The preliminary edition of this book has been used as a text for a nonlinear circuit analysis course at Purdue University. This course consists of three lecture hours per week and is taken by all electrical engineering undergraduate students in the second semester of the sophomore year. Concurrently, the students take a course in linear active circuits. In addition to these two courses, the students take a six-hour-per-week laboratory course in which they are introduced to a wide range of devices, including the latest commercially available. Most of the practical nonlinear electronic circuits presented in this book (electronic switches, gates, logic circuits, multivibrators, time-base generators, etc.) are built and verified by the students in this laboratory course. Our experience has shown this combination of courses to be very successful.

This book could not have been written without the generous support of various organizations and individuals. In fact, most of the original results included in this book are the product of the author's research under the sponsorships of the National Science Foundation and the Army, Navy, and Air Force Joint Services Electronics Program. I am particularly grateful to Dr. J. C. Hancock of Purdue University, whose leadership in electrical engineering education has been instrumental in a complete revision of the Purdue electrical engineering curriculum. This book is a by-product of that revision. I am indebted to Dr. R. A. Rohrer of the University of California (Berkeley) and Dr. M. E. Van Valkenburg of Princeton University for their constant encouragement and advice. I also wish to acknowledge a number of my colleagues who offered valuable suggestions during the evolution of this book. They are Dr. C. A. Desoer and Dr. E. S. Kuh of the University of California (Berkeley), Dr. Y-L Kuo, Dr. B. J. Leon, Dr. D. A. Landgrebe, Dr. P-M Lin, Dr. C. D. McGillem, and Dr. L. F. Silva, all of Purdue University. A great number of students have helped in one way or another throughout the course of this work. Among them, I wish to single out Mr. J. C. Evans, Mr. D. A.

Perreault, and Mr. K. A. Stromsmoe for their many critical suggestions in transforming the preliminary edition to the present form. I would also wish to thank Mr. Michael Elia of the McGraw-Hill Book Company for his valuable editorial help. But most of all, this book would never have been finished without the infinite patience and assistance of my wife, to whom this book is dedicated.

Leon O. Chua

CONTENTS

Preface, vii

List of Reference Tables, xxv

Glossary of Symbols and Abbreviations, xxvi

PART 1 FOUNDATIONS OF NONLINEAR NETWORK THEORY

1 **TWO-TERMINAL NETWORK ELEMENTS,** 3

- 1-1 *Review of Basic Physical Variables in Network Theory,* 3
- 1-2 *The Simultaneity Postulate in Lumped-network Theory,* 5
- 1-3 *Significance of the Reference Current Direction and the Reference Voltage Polarity,* 7
- 1-4 *Independent Sources,* 9
- 1-5 *Characterization of a Two-terminal Black Box,* 11
  - 1-5-1 *A Mechanical Black-box Analogy,* 12
  - 1-5-2 *Static Characteristics of a Two-terminal Black Box,* 15
- 1-6 *Two-terminal Resistors,* 17
  - 1-6-1 *Linear Resistors,* 17
  - 1-6-2 *Nonlinear Resistors,* 18
  - 1-6-3 *Classification of* v-i *Curves,* 19
  - 1-6-4 v-i *Curves of DC Sources and Ideal Diodes,* 23
  - 1-6-5 *Some Practical Applications of Two-terminal Nonlinear Resistors,* 24
- 1-7 *Two-terminal Capacitors,* 30
  - 1-7-1 *Linear Capacitors,* 30
  - 1-7-2 *Nonlinear Capacitors,* 31
  - 1-7-3 *Some Practical Applications of Two-terminal Nonlinear Capacitors,* 32
- 1-8 *Two-terminal Inductors,* 35
  - 1-8-1 *Linear Inductors,* 35
  - 1-8-2 *Nonlinear Inductors,* 36
  - 1-8-3 *Some Practical Applications of Two-terminal Nonlinear Inductors,* 37
- 1-9 *Energy and Power,* 38
- 1-10 *Time-varying Elements,* 47

1-11 *Concepts of Modeling,* 51
1-12 *Summary,* 53

2 **CONTROLLED ELEMENTS,** 62

- 2-1 *Two-terminal Controlled Elements,* 62
  - 2-1-1 *Two-terminal Controlled Resistors,* 62
  - 2-1-2 *Two-terminal Controlled Capacitors,* 67
  - 2-1-3 *Two-terminal Controlled Inductors,* 67
- 2-2 *Practical Applications of Controlled Elements,* 68
- 2-3 *Controlled Sources,* 70
  - 2-3-1 *Transducers,* 71
  - 2-3-2 *Sources Controlled by Electrical Variables,* 72
- 2-4 *A Basic Composition Technique,* 75
- 2-5 *Summary,* 78

3 **MULTITERMINAL ELEMENTS,** 84

- 3-1 *Characterization of a Multiterminal Black Box,* 84
- 3-2 *Three-terminal Resistors,* 86
  - 3-2-1 *Forms of Representation,* 89
  - 3-2-2 *Graphical Transformation of Representations,* 95
  - 3-2-3 *Transformation to Another Common Terminal,* 97
  - 3-2-4 *Some Practical Three-terminal Resistors,* 98
- 3-3 *Three-terminal Capacitors,* 103
- 3-4 *Three-terminal Inductors,* 104
- 3-5 *Multiterminal Elements,* 106
  - 3-5-1 *Multiterminal Resistors Operating as Three-terminal Controlled Resistors,* 107
  - 3-5-2 *Multiterminal Elements with Prescribed Constraints,* 108
- 3-6 *Some Practical Applications of Operational Amplifiers,* 115
  - 3-6-1 *Operational Amplifier Used as a Nonlinear Element,* 115
  - 3-6-2 *Operational Amplifier Used as a Linear Element,* 117
- 3-7 *Scalors, Rotators, and Reflectors: A Class of Useful Two-port Resistors,* 128
  - 3-7-1 *Scalors,* 130
  - 3-7-2 *Rotators,* 131
  - 3-7-3 *Reflectors,* 134
- 3-8 *Mutator: The Chameleon Black Box,* 138
  - 3-8-1 R-L *Mutators,* 138
  - 3-8-2 R-C *Mutators,* 140
  - 3-8-3 C-L *Mutators,* 142
- 3-9 *Summary,* 144

4 **EQUATIONS OF MOTION,** 152

- 4-1 *Classification of Nonlinear Networks,* 152
- 4-2 *Laws of Elements and Laws of Interconnection,* 153
  - 4-2-1 *Equations of Motion Pertaining to the Laws of Elements,* 155
  - 4-2-2 *Equations of Motion Pertaining to the Laws of Interconnection,* 155

4-3 *Introduction to Network Topology,* 157
  4-3-1 *Criteria for Writing Independent KVL Equations,* 160
  4-3-2 *Criteria for Writing Independent KCL Equations,* 163
4-4 *Equations of Motion for Resistive Networks,* 166
  4-4-1 *Networks Containing Only Two-terminal Resistors and Independent Sources,* 166
  4-4-2 *Networks Containing Multiterminal Resistors and Controlled Sources,* 173
4-5 *Practical Methods for Solving Nonlinear Functional Equations,* 179
  4-5-1 *The Fixed-point Concept,* 180
  4-5-2 *The Newton-Raphson Method,* 185
  4-5-3 *Equations with Many Unknowns,* 187
4-6 *Equations of Motion for Dynamic Networks,* 188
  4-6-1 *Selection of State Variables of Normal-form Equations,* 192
  4-6-2 *Autonomous and Nonautonomous Networks,* 195
4-7 *Practical Methods for Solving Nonlinear Differential Equations,* 196
  4-7-1 *Euler Algorithm for One Differential Equation,* 196
  4-7-2 *Euler Algorithm for Two or More Differential Equations,* 199
4-8 *Principles of Duality,* 201
  4-8-1 *Duality Relationships from the Laws of Elements,* 203
  4-8-2 *Duality Relationships from the Laws of Interconnection,* 204
  4-8-3 *Algorithm for Drawing the Dual of a Planar Network,* 209
4-9 *Summary,* 213

PART 2 RESISTIVE NONLINEAR NETWORKS

5 THREE BASIC CONCEPTS OF RESISTIVE NONLINEAR NETWORKS, 225

5-1 *The Operating-point Concept,* 225
5-2 *Concepts of Driving-point and Transfer-characteristic Plots,* 228
  5-2-1 *The Driving-point Characteristic Plot (DP Plot),* 231
  5-2-2 *The Transfer-characteristic Plot (TC Plot),* 232
5-3 *Some Practical Examples,* 236
  5-3-1 *Isolation Block for Integrated Circuits,* 236
  5-3-2 *A Square-law Transfer-characteristic Plot,* 238
5-4 *The Power-Transfer Plot,* 241
5-5 *The Operating-point Paradox,* 243
5-6 *The Three Fundamental Theorems of Resistive Networks,* 244
5-7 *Summary,* 246

6 GRAPHICAL ANALYSIS OF RESISTIVE NONLINEAR NETWORKS, 253

6-1 *A Bird's-eye View,* 253
6-2 *Graphical Determination of the Operating Point,* 254

    - 6-2-1 *Basic Network Configuration 1, Interconnection between 2 Two-terminal Resistors,* 255
    - 6-2-2 *Basic Network Configuration 2, Interconnection between 2 Two-terminal Resistors and a Three-terminal Resistor,* 260
- 6-3 *Graphical Determination of DP Plots of Series-Parallel Networks,* 266
    - 6-3-1 *The Series-combination Technique,* 267
    - 6-3-2 *The Parallel-combination Technique,* 271
    - 6-3-3 *Combination of Series-parallel Techniques,* 272
- 6-4 *Some Practical Applications of DP Plot and Operating-point Concepts,* 273
    - 6-4-1 *Concave and Convex Resistors,* 273
    - 6-4-2 *Elements with a Horizontal Segment,* 276
    - 6-4-3 *Elements with a Vertical Segment,* 277
    - 6-4-4 *Multithreshold Elements,* 277
    - 6-4-5 *Automatic Sorting Circuit,* 280
    - 6-4-6 *Automatic Comparison and Null-detection Circuit,* 282
- 6-5 *Graphical Determination of TC Plots of Series-parallel Networks,* 286
    - 6-5-1 *Nonlinear Voltage Divider,* 286
    - 6-5-2 *Nonlinear Current Divider,* 290
    - 6-5-3 *Series-parallel Nonlinear Networks,* 290
- 6-6 *Some Practical Applications of TC Plots,* 294
    - 6-6-1 *Half-wave Rectifier,* 294
    - 6-6-2 *Voltage Limiter,* 296
    - 6-6-3 *Pulse Compressor,* 297
    - 6-6-4 *Vertical-pulse Generator,* 298
- 6-7 *DP Plot and TC Plot of Networks Containing Three-terminal Resistors,* 301
    - 6-7-1 *The Template–Double Load-line Method,* 303
    - 6-7-2 *Some Practical Circuits,* 306
- 6-8 *Composite Characteristics of Three-terminal Resistive Black Boxes,* 311
    - 6-8-1 *Four Basic Three-terminal Black Boxes,* 311
    - 6-8-2 *Cascade Connection of Basic Three-terminal Black Boxes,* 313
- 6-9 *Summary,* 316

7 PRINCIPLES OF EQUIVALENCE AND SYMMETRY, 325

- 7-1 *Some Shortcuts in Nonlinear Network Analysis,* 325
- 7-2 *Definition of Equivalent Networks,* 325
- 7-3 *Equivalence Based on Identical DP Plots,* 327
- 7-4 *Equivalence Based on Identical Operating Points,* 330
- 7-5 *The Shifting Theorems,* 332
- 7-6 *Graphical Analysis of Non-series-parallel Networks,* 334
- 7-7 *Principles and Applications of Symmetry,* 342
    - 7-7-1 *Some Applications of Symmetry to the Operating-point Problem,* 343
    - 7-7-2 *Some Applications of Symmetry to the Determination of DP Plots and TC Plots,* 348
- 7-8 *Complementary Networks,* 353

- 7-9 *Some Practical Circuits,* 355
  - 7-9-1 *DC Amplifier,* 355
  - 7-9-2 *Complementary Symmetric Amplifier,* 356
  - 7-9-3 *Push-pull Amplifier,* 357
- 7-10 *Summary,* 361

8 GRAPHICAL SYNTHESIS OF RESISTIVE NONLINEAR NETWORKS, 368

- 8-1 *What Is Network Synthesis?* 368
- 8-2 *Synthesis of Operating Points,* 369
  - 8-2-1 *Biasing a Two-terminal Nonlinear Resistor,* 369
  - 8-2-2 *Biasing a Three-terminal Nonlinear Resistor,* 374
  - 8-2-3 *Sensitivity Consideration in Three-terminal Biasing Circuits,* 382
- 8-3 *Synthesis of DP Plots,* 386
  - 8-3-1 *Synthesis of Monotonic DP Plots,* 388
  - 8-3-2 *Synthesis of Nonmonotonic DP Plots,* 396
- 8-4 *Synthesis of TC Plots,* 401
  - 8-4-1 *TC Plot Synthesis by a Nonlinear Voltage- or Current-divider Network,* 402
  - 8-4-2 *TC Plot Synthesis by Operational Amplifier Circuits,* 405
- 8-5 *Synthesis of Jointly Prescribed DP Plot, TC Plot, and Load* v-i *Curve,* 408
- 8-6 *Summary,* 413

9 SYNTHESIS OF DC-RESISTIVE NONLINEAR FUNCTIONAL NETWORKS: THE BLACK-BOX APPROACH, 422

- 9-1 *Basic Philosophy,* 422
- 9-2 *Functional Black Box for Waveform Generation,* 424
  - 9-2-1 *Generation of Periodic Waveforms,* 425
  - 9-2-2 *Generation of Aperiodic Waveforms,* 426
- 9-3 *Functional Black Box for Waveshaping,* 427
  - 9-3-1 *Clipping Networks,* 427
  - 9-3-2 *Analog-to-digital Converters,* 431
- 9-4 *Functional Black Box for Compensation,* 433
- 9-5 *Functional Black Box for Regulation,* 437
- 9-6 *Summary,* 442

10 SYNTHESIS OF AC-RESISTIVE NONLINEAR FUNCTIONAL NETWORKS: THE BLACK-BOX APPROACH, 449

- 10-1 *Characteristics of Electronically Controlled Switches,* 449
- 10-2 *Synthesis of Single-controlled Undirectional Switches,* 452
  - 10-2-1 *A High-switching-sensitivity Transistor Switch,* 454
  - 10-2-2 *A High-switching-sensitivity SCR Latching Switch,* 456
- 10-3 *Synthesis of Single-controlled Bidirectional Switches,* 458
- 10-4 *Synthesis of Choppers and Amplitude Modulators,* 461
  - 10-4-1 *Practical Choppers,* 462
  - 10-4-2 *Practical Amplitude Modulators,* 463

- 10-5 *Synthesis of Multicontrolled Electronic Switches,* 465
  - 10-5-1 *Synthesis of a Coincidence Gate,* 466
  - 10-5-2 *Synthesis of Matrix Transmission Gates,* 468
- 10-6 *Synthesis of Logic Circuit Building Blocks,* 469
  - 10-6-1 *Synthesis of AND Gates,* 472
  - 10-6-2 *Synthesis of OR Gates,* 474
  - 10-6-3 *Synthesis of NAND, NOR, and NOT Gates,* 476
- 10-7 *Summary,* 479

11 SYNTHESIS OF RESISTIVE NONLINEAR NETWORK MODELS, 489

- 11-1 *What Is a Network Model?* 489
- 11-2 *Principles of Model Making,* 492
- 11-3 *Synthesis of Global Models for Nonlinear Controlled Resistors,* 494
- 11-4 *Synthesis of Global Models for Nonlinear Three-terminal Resistors,* 497
  - 11-4-1 *A Basic T-network Model,* 498
  - 11-4-2 *A Basic* π*-network Model,* 501
- 11-5 *Some Useful Techniques for Modifying Models,* 503
  - 11-5-1 *Constraining* v-i *Curves to Half-planes,* 503
  - 11-5-2 *Constraining* v-i *Curves to Quadrants,* 505
  - 11-5-3 *Constraining* v-i *Curves to the First and Third Quadrants,* 506
- 11-6 *Models of Practical Three-terminal Resistors,* 508
  - 11-6-1 *Vacuum Triode Models,* 508
  - 11-6-2 *Vacuum Pentode Models,* 510
  - 11-6-3 n-p-n *Transistor Models,* 512
  - 11-6-4 p-n-p *Transistor Models,* 517
  - 11-6-5 n-*channel FET Models,* 517
  - 11-6-6 p-*channel FET Models,* 517
- 11-7 *Summary,* 520

12 ITERATIVE PIECEWISE-LINEAR ANALYSIS AND SYNTHESIS OF RESISTIVE NONLINEAR NETWORKS, 526

- 12-1 *Basic Philosophy of the Iterative Piecewise-linear Method,* 526
- 12-2 *Determination of Operating Points,* 530
- 12-3 *Determination of DP Plots,* 539
- 12-4 *Determination of TC Plots,* 553
- 12-5 *Computer Analysis of Resistive Nonlinear Networks,* 558
  - 12-5-1 *Semiautomatic General Analysis Program,* 558
  - 12-5-2 *Completely Automatic Meca Program,* 559
- 12-6 *TC Plot Synthesis by the Piecewise-linear Method,* 562
  - 12-6-1 *TC Plot Synthesis by Nonlinear Voltage Dividers,* 562
  - 12-6-2 *Jointly Prescribed DP Plot, TC Plot, and Load* v-i *Curve Synthesis,* 565
- 12-7 *Summary,* 573

PART 3 DYNAMIC NONLINEAR NETWORKS

13 BASIC CONCEPTS OF DYNAMIC NONLINEAR NETWORKS, 583

13-1 *Classification of Dynamic Nonlinear Networks,* 583
  13-1-1 *Basis of Classification,* 583
  13-1-2 *Definition of Order of Complexity,* 585
13-2 *Order of Complexity of Dynamic Networks,* 586
  13-2-1 *Significance of the Initial Condition,* 586
  13-2-2 *Determination of Order of Complexity by Inspection,* 589
  13-2-3 *Some Subtle Points Concerning Initial Conditions,* 592
13-3 *Basic Principles for Analyzing Dynamic Networks,* 593
13-4 *Deriving Normal-form Equations from the Characteristics of the* N*-port Subnetwork,* 595
  13-4-1 *Normal-form Equations for First-order Networks,* 595
  13-4-2 *Normal-form Equations for Second-order Networks,* 599
  13-4-3 *Normal-form Equations for* n*th-order Networks,* 604
13-5 *Trajectory of Solutions in the State Space,* 606
13-6 *Equilibrium States of Autonomous Networks,* 613
  13-6-1 *Equilibrium States of a Switching Circuit: A Numerical Example,* 617
  13-6-2 *Relationship between Trajectories and Equilibrium States,* 620
13-7 *Stability of Equilibrium States,* 622
  13-7-1 *Stability Definition for Second-order Networks,* 624
  13-7-2 *Stability Definition for* n*th-order Networks,* 626
13-8 *Summary,* 626

14 ANALYSIS OF AUTONOMOUS FIRST-ORDER NONLINEAR NETWORKS, 636

14-1 *Basic Philosophy and Approach,* 636
14-2 *Equilibrium States and Stability Criteria,* 637
14-3 *Stability Criteria for Two Common Circuit Configurations,* 640
14-4 *Time Scaling the Trajectory of Solutions,* 644
14-5 *Realistic and Incomplete Models,* 649
14-6 *Jump Phenomenon and Oscillatory Solution in First-order Networks,* 652
14-7 *Incomplete Models Requiring Additional Parasitic Elements,* 656
14-8 *The Piecewise-linear Approach,* 659
  14-8-1 *Networks Containing a Linear Energy-storage Element across a Nonlinear Resistive Network Characterized by Piecewise-linear Segments with Finite Nonzero Slopes,* 660

- 14-8-2 *Networks Containing a Linear Energy-storage Element across a Nonlinear Resistive Network Characterized by Piecewise-linear Segments with Arbitrary Slopes,* 682
- 14-8-3 *Networks Containing a Piecewise-linear Energy-storage Element across a Nonlinear Resistive Network Characterized by Piecewise-linear Segments with Arbitrary Slopes,* 686
- 14-9 *Summary,* 689

15 ANALYSIS OF FIRST-ORDER NONLINEAR SWITCHING NETWORKS, 697

- 15-1 *What Is a First-order Switching Network?* 697
- 15-2 *Analysis of First-order Linear Switching Networks,* 698
- 15-3 *Analysis of First-order Nonlinear Switching Networks,* 708
- 15-4 *Analysis of Autonomous First-order Networks Containing One Nonlinear Inductor or Capacitor,* 711
- 15-5 *Summary,* 718

16 SYNTHESIS OF FIRST-ORDER MULTIVIBRATORS: THE BLACK-BOX APPROACH, 725

- 16-1 *What Is a First-order Multivibrator?* 725
- 16-2 *Astable (Free-running) Multivibrators,* 726
- 16-3 *Monostable (Single-shot) Multivibrators,* 730
  - 16-3-1 *Some Applications of Monostable Multivibrators,* 733
  - 16-3-2 *Minimum Triggering-pulse-amplitude Condition,* 735
  - 16-3-3 *Minimum Triggering-pulse-width Condition,* 736
- 16-4 *Bistable (Double-shot, Flip-flop) Multivibrators,* 737
- 16-5 *Summary,* 746

17 SYNTHESIS OF FIRST-ORDER TIME-BASE GENERATORS: THE BLACK-BOX APPROACH, 753

- 17-1 *What Is a Time-base Generator?* 753
- 17-2 *Synthesis of Free-running Time-base Generators,* 759
- 17-3 *Synchronization of Free-running Time-base Generators,* 763
- 17-4 *Synthesis of Triggered Time-base Generators,* 768
- 17-5 *Summary,* 774

18 ANALYSIS OF NONAUTONOMOUS FIRST-ORDER NONLINEAR NETWORKS, 778

- 18-1 *Motivation and Strategies,* 778
- 18-2 *Nonautonomous First-order Linear Network Analysis,* 779
  - 18-2-1 *Zero-state Response,* 779
  - 18-2-2 *Complete Response,* 781
  - 18-2-3 *Very Large and Very Small Time-constant Networks,* 783
- 18-3 *Nonautonomous First-order Piecewise-linear Network Analysis,* 786

    - 18-3-1 *One Nonlinear Resistor Case,* 787
    - 18-3-2 M *Nonlinear Resistor Case,* 790
- 18-4 *The Constant-slope Network Approach,* 793
    - 18-4-1 *Methods for Drawing the Vector Field,* 794
    - 18-4-2 *The Concept of Constant-slope Networks,* 797
    - 18-4-3 *Some Illustrative Examples,* 800
- 18-5 *Some Practical Applications of Nonautonomous First-order Nonlinear Networks,* 807
    - 18-5-1 *The Monotonic Charging Property and Its Applications,* 807
    - 18-5-2 *The Pulse-smoothing Property,* 811
    - 18-5-3 *The Pulse-sharpening Property,* 817
- 18-6 *Summary,* 819

19 ANALYSIS OF AUTONOMOUS SECOND-ORDER NONLINEAR NETWORKS, 823

- 19-1 *The Phase-plane Technique,* 823
    - 19-1-1 *Isoclines and Phase Portraits,* 823
    - 19-1-2 *Time Scaling the Trajectories,* 828
- 19-2 *The Isocline Network Method,* 833
    - 19-2-1 *The Isocline Network,* 833
    - 19-2-2 *Some Practical Networks,* 836
- 19-3 *From Qualitative to Quantitative Analysis,* 842
- 19-4 *Analytical Solutions of Autonomous Second-order Linear Networks,* 843
- 19-5 *Piecewise-linear Analysis of Autonomous Second-order Nonlinear Networks,* 847
- 19-6 *Behavior of Trajectories near an Equilibrium State,* 859
    - 19-6-1 *Local Behavior of Trajectories in the Aperiodic Case,* 861
    - 19-6-2 *Local Behavior of Trajectories in the Spiral Case,* 867
    - 19-6-3 *Summary of Singular-point Classifications,* 868
    - 19-6-4 *The Phase Portrait,* 871
- 19-7 *Summary,* 873

20 ANALYSIS OF nTH-ORDER NONLINEAR NETWORKS, 880

- 20-1 *Beyond the Second-order Autonomous Network Barrier,* 880
- 20-2 *The Numerical-integration Approach,* 881
- 20-3 *The Piecewise-linear Approach,* 884
- 20-4 *The Heuristic Approach,* 885
- 20-5 *The Qualitative Approach,* 890
- 20-6 *The Approximate Analytical Approach,* 895
- 20-7 *Summary,* 895

APPENDIX A MATHEMATICAL REPRESENTATION OF NONLINEAR NETWORK ELEMENTS, 898

- A-1 *Relations and Their Inverses,* 898
- A-2 *Functions and Their Inverses,* 901
- A-3 *Mathematical Representation of Functions,* 904

A-4 *Mathematical Representation of Multivalued Functions,* 909
A-5 *Parametric Representation of Nonlinear Network Elements,* 913

APPENDIX B GRAPHICAL TECHNIQUES FOR BASIC MATHEMATICAL OPERATIONS, 919

B-1 *Graphical Addition and Subtraction,* 920
B-2 *Graphical Multiplication and Division,* 920
B-3 *Graphical Composition,* 921
B-4 *Graphical Elimination,* 923
B-5 *Graphical Differentiation,* 924
B-6 *Graphical Integration,* 926

APPENDIX C SCHEMATIC DIAGRAMS OF SOME BASIC CIRCUITS, 929

C-1 *Circuits for Tracing* v-i, v-q, *and* i-φ *Curves,* 929
C-2 *Circuits for Simulating Controlled Sources,* 932
C-3 *Circuits for Simulating Negative-impedance Converters,* 937
C-4 *Circuits for Simulating Scalors,* 938
C-5 *Circuits for Simulating Rotators,* 939
C-6 *Circuits for Simulating Reflectors,* 939
C-7 *Circuits for Simulating Mutators,* 940

APPENDIX D CHARACTERISTIC CURVES OF TYPICAL DEVICES, 943

APPENDIX E SELECTED LIST OF REFERENCES, 958

E-1 *References on Basic Circuit Theory,* 958
E-2 *References on Basic Electronic Circuits and Devices,* 959
E-3 *References on Linear Active Networks,* 959
E-4 *References on Nonlinear Electronic Circuits,* 960
E-5 *References on Nonlinear Network Elements,* 960
E-6 *References on Integrated Circuits,* 963
E-7 *References on Device Modeling,* 964
E-8 *References on the Formulation of Equations of Motion of Nonlinear Networks,* 964
E-9 *References on Numerical Techniques,* 965
E-10 *References on Computer-aided Analysis,* 965
E-11 *References on General Nonlinear Network Properties and Theorems,* 966
E-12 *References on Approximation Techniques,* 968
E-13 *References on Qualitative Methods of Analysis,* 968
E-14 *References on Analytical Techniques,* 969

Index of Theorems, 973

Subject Index, 975

LIST OF REFERENCE TABLES

1-1 *Practical two-terminal resistors,* 20

1-2 *Instantaneous power, energy, and average power flow in a nonlinear resistor,* 41

2-1 *Practical two-terminal controlled resistors,* 64

3-1 *A list of common three-terminal resistors,* 100

4-6 *Common dual quantities,* 210

8-1 *Design equations for three common biasing circuits,* 380

8-2 *Basic building blocks for synthesis,* 387

14-1 *General properties characterizing the exponential solution of first-order networks,* 665

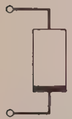
16-1 *Properties of first-order multivibrator,* 746

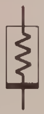
19-1 *Summary of solutions of autonomous second-order linear networks and their typical trajectories,* 848

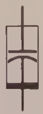
19-5 *Summary of singular-point classifications,* 868

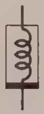
GLOSSARY OF SYMBOLS AND ABBREVIATIONS

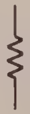
TWO-TERMINAL ELEMENTS

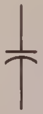

A hypothetical two-terminal element

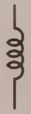

Nonlinear resistor

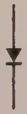

Nonlinear capacitor

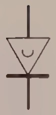

Nonlinear inductor

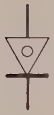

Linear resistor

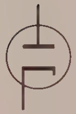

Linear capacitor

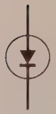

Linear inductor

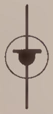

Ideal diode

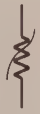

Concave resistor

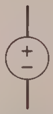

Convex resistor

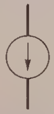

Vacuum diode

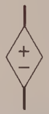

Junction diode

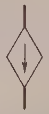

Back diode

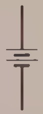

Varistor

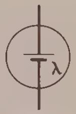

Independent voltage source

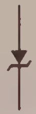

Independent current source

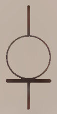

Controlled voltage source

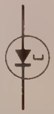

Controlled current source


Battery

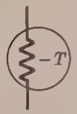

Solar cell

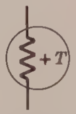

Zener diode

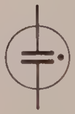

Constant-current diode

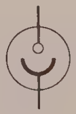

Tunnel diode

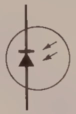
CONTROLLED ELEMENTS


Negative-temperature-coefficient (NTC) thermistor

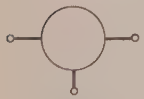

Positive-temperature-coefficient (PTC) thermistor

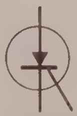

Glow tube (neon bulb)

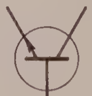

Phototube

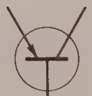

Photodiode

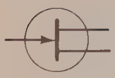
THREE-TERMINAL ELEMENTS

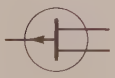

A hypothetical three-terminal element

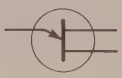

Silicon-controlled rectifier (SCR)

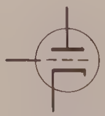

n-p-n transistor


p-n-p transistor

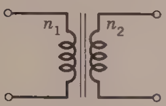

N-channel FET


P-channel FET

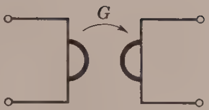

Unijunction transistor


Triode

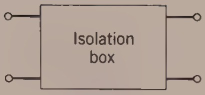
TWO-PORT ELEMENTS



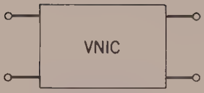
Ideal transformer



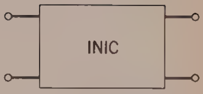
Gyrator



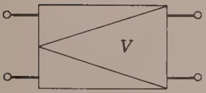
Isolation box



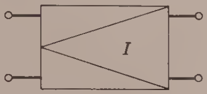
Voltage-inversion negative-impedance converter



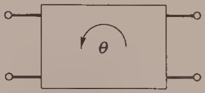
Current-inversion negative-impedance converter



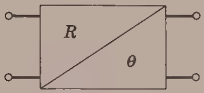
Voltage scalor



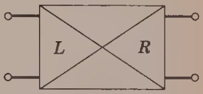
Current scalor



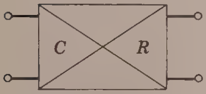
Rotator



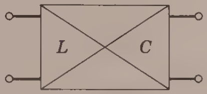
Reflector



R-L mutator

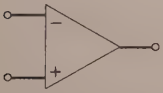


R-C mutator

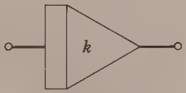


C-L mutator

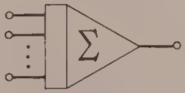
N-TERMINAL ELEMENTS



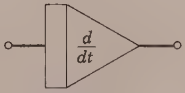
Operational amplifier



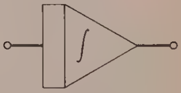
Scaling circuit



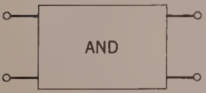
Summing circuit



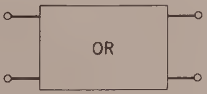
Differentiator



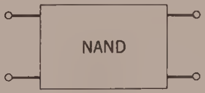
Integrator



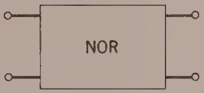
AND gate



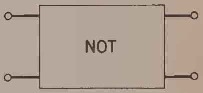
OR gate



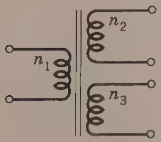
NAND gate



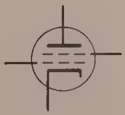
NOR gate



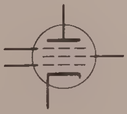
NOT gate



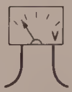
Three-port transformer



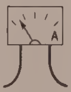
Tetrode



Pentode



Voltmeter



Ammeter

ABBREVIATIONS

DP plot – Driving-point-characteristic plot
TC plot – Transfer-characteristic plot
AM – Amplitude modulation
FET – Field-effect transistor
VNIC – Voltage-inversion negative-impedance converter
INIC – Current-inversion negative-impedance converter
SCR – Silicon-controlled rectifier

INTRODUCTION TO

NONLINEAR NETWORK THEORY

part 1

FOUNDATIONS OF NONLINEAR NETWORK THEORY

1 TWO-TERMINAL NETWORK ELEMENTS

1-1 REVIEW OF BASIC PHYSICAL VARIABLES IN NETWORK THEORY

The advent of electrical science occurred with the discovery of the phenomenon that dry substances such as amber or rubber tend to repel or attract each other upon being rubbed by different materials such as silk or fur. This phenomenon was first explained by postulating the existence of a certain basic electrical quantity called the "electric charge" q, which may be either positive or negative, and which has the property that like charges exert a force of repulsion and unlike charges exert a force of attraction between each other. The quantity "charge" remains the most basic electrical quantity today, and its existence can now be explained by the atomic theory: a body is "charged" whenever there is an excess of the positive charges in the nucleus over the negatively charged electrons and vice versa. The practical unit of charge called the *coulomb* has been defined to be equivalent to the total charge possessed by 6.24×10^{18} electrons. The quantity of charge possessed by a body can be measured by various instruments such as the electroscope.

Since charged bodies exert forces on one another, energy or work is involved whenever one charged body is moved in the vicinity of another charged body. Hence if w is the work done by moving a charge q from point j to point k (assuming w is independent of the path taken),[1] then the potential difference, or voltage, between these points is defined as the work per unit charge; that is,

$$v_{jk} = \frac{w}{q} \tag{1-1}$$

---

[1] This assumption is only approximately satisfied in practice. The study of the conditions under which this assumption is valid belongs to a course in electromagnetic field theory. However, as far as network theory is concerned, the above assumption is automatically implied. Very roughly speaking, the above assumption is valid when the frequency of the signal is "not too high," that is, when the wavelength of the signals is long compared with the dimension of the physical network.

Observe that the magnitude of the charge is arbitrary; only the ratio between work and charge is important. Hence, the incremental work dw required to move an incremental test charge dq from point j to point k must also satisfy Eq. (1-1); thus

$$v_{jk} = \frac{dw}{dq} \tag{1-2}$$

When there is no possibility of confusion, we can delete the subscripts j and k and express the voltage simply as v. The practical unit of voltage is called the *volt*. The voltage v between two points can be measured by a voltmeter.

Charges can be caused to flow from one charged body into another by connecting a conducting wire between the two bodies. In 1819, Hans Christian Oersted discovered that the flow of charge through a wire produced a force on a compass needle in the vicinity of the wire and that force was proportional to the rate of flow of charge. Since the force on a compass needle can be easily determined by noting the deflection of the needle, the quantity "rate of flow of the charge" becomes very useful, and it has been given the name current, i. By definition,

$$i = \frac{dq}{dt} \tag{1-3}$$

The practical unit of current is the *ampere;* i.e., one ampere represents a charge flowing at a rate of one coulomb per second. The current i can be measured by an ammeter.

The deflection of a magnetic compass needle caused by the flow of charge, or current, in a conductor indicates that current produces a magnetic effect. This effect can be explained by the generation of a magnetic flux λ by the current. If the conductor is wound into a coil of n turns, then by defining $\varphi = n\lambda$ to be the flux linkage, Faraday discovered that the voltage between the two terminals of the coil is given simply by

$$v = \frac{d\varphi}{dt} \tag{1-4}$$

The practical unit of the flux linkage φ is called the *weber*. Flux linkage can be measured by a fluxmeter.

If we multiply together the left and right sides of Eqs. (1-2) and (1-3), we obtain

$$vi = \frac{dw}{dq}\frac{dq}{dt} = \frac{dw}{dt} \tag{1-5}$$

Since w represents energy, dw/dt represents the rate of change of energy, or the power p; hence

$$p(t) = v(t)i(t) \tag{1-6}$$

Summarizing, therefore, we find that the six basic electrical quantities of interest in network theory are the charge q, the voltage v, the current i, the flux linkage φ, the power p, and the energy w. The universal relationships between these quantities at any time t are

$$i(t) = \frac{dq(t)}{dt} \tag{1-7}$$

$$v(t) = \frac{d\varphi(t)}{dt} \tag{1-8}$$

$$p(t) = v(t)i(t) = \frac{dw(t)}{dt} \tag{1-9}$$

$$w(t) = \int_{-\infty}^{t} p(\tau)\, d\tau = \int_{-\infty}^{t} v(\tau)i(\tau)\, d\tau \tag{1-10}$$

$$q(t) = \int_{-\infty}^{t} i(\tau)\, d\tau \tag{1-11}$$

$$\varphi(t) = \int_{-\infty}^{t} v(\tau)\, d\tau \tag{1-12}$$

1-2 THE SIMULTANEITY POSTULATE IN LUMPED-NETWORK THEORY

The six basic electrical variables related by Eqs. (1-7) to (1-12) are assumed to be functions of only one independent variable, namely, the *time* of measurement t. Actually, to be exact, we must introduce another independent variable for specifying the relative location of the various terminals at which these electrical quantities are to be measured. This is the variable *length*, or *dimension*, in centimeters. The necessity for introducing this variable is due to the fact that it takes a finite amount of time for electrons to move from one point to another. For example, if we apply a voltage $v_s(t)$ across one end of a 30-cm lossless transmission line as shown in Fig. 1-1*a*, it will take 1 nsec ($30\text{ cm}/3 \times 10^{10}\text{ cm/sec} = 10^{-9}$ sec) for the signal to arrive at the other end ($x = 30$ cm).[1] If

---

[1] For simplicity, we assume the electrons traverse down the line at the velocity of light. The actual electron velocity will, of course, depend on the characteristics of the transmission line.

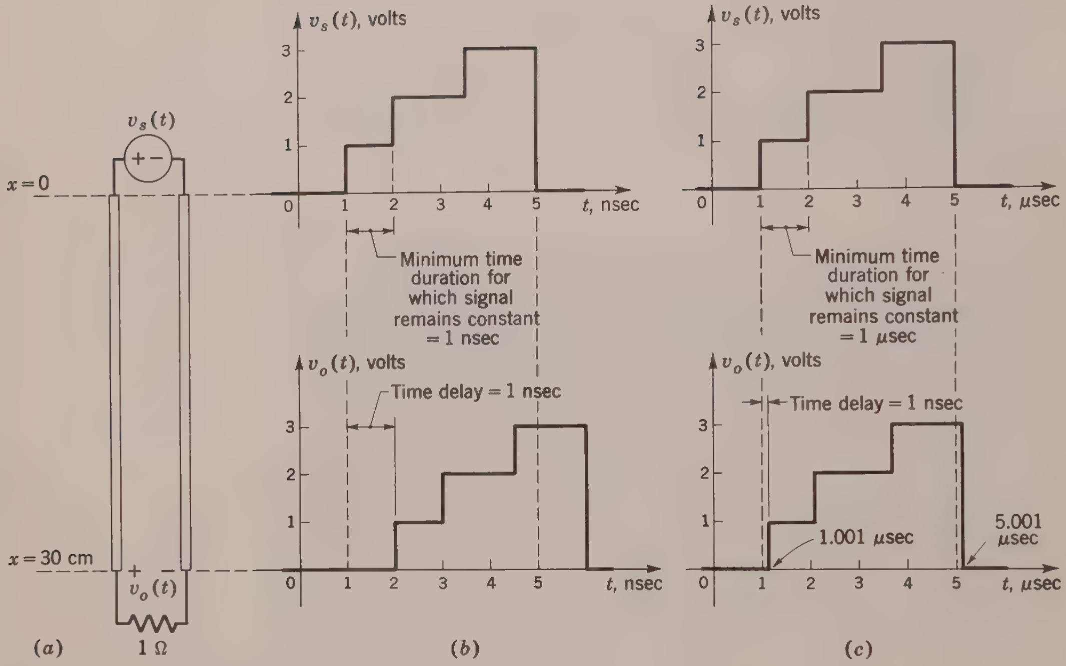


Fig. 1-1. The length of the transmission line introduces a time delay which is significant in (*b*) but may be neglected in (*c*).

the duration of time for which the signal level remains relatively unchanged is of the same order of magnitude (say, 2 nsec), then the time delay of the transmission line cannot be neglected. This is easily seen by comparing the signals $v_s(t)$ and $v_o(t)$ as shown in Fig. 1-1*b*. On the other hand, if the signal level does not change rapidly (relative to the time delay) as in Fig. 1-1*c*, then the time delay is insignificant and may therefore be neglected. Under this assumption, the output signal $v_o(t)$ may be considered to appear at the same instant as the input signal $v_s(t)$. This is equivalent to the assumption that the length of the transmission line is insignificant. In other words, the line can be *lumped* as one point so that *the current entering one terminal of a terminal pair appears instantaneously at the other terminal.* We will refer to this assumption as the *simultaneity postulate.*

The simultaneity postulate is a fundamental assumption in lumped-network theory that applies not only to transmission lines but also to all two-terminal black boxes considered in this book.[1] This postulate is valid whenever the physical dimension of each device inside the black box is small so that the time delay it introduces is insignificant compared with the minimum time duration for which the signals remain relatively constant. For periodic

---

[1] Networks which do not satisfy the simultaneity postulate are said to be distributed. The study of distributed networks belongs to a course in electromagnetic field theory.

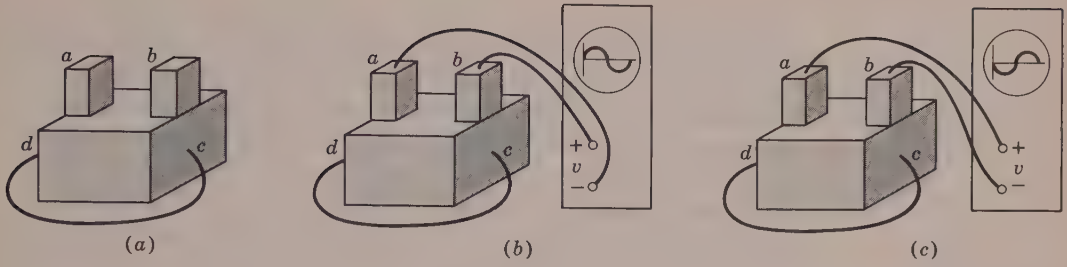


Fig. 1-2. An experiment demonstrating that regardless of which terminal of the black box is chosen to be positive, the actual voltage across terminals *a-b* can be unambiguously specified for all time.

signals, the reciprocal of the frequency is a good measure of this minimum time duration. Hence, roughly speaking, the higher the operating frequency, the smaller must be the device's physical dimension in order for the simultaneity postulate to be satisfied.[1] Fortunately, most nonlinear electronic circuits of interest do satisfy the simultaneity postulate. This is especially true with *integrated circuits* where the components are becoming so small that they can be seen only with the aid of a microscope.

1-3 SIGNIFICANCE OF THE REFERENCE CURRENT DIRECTION AND THE REFERENCE VOLTAGE POLARITY

One of the most basic concepts in physical science is that any physical quantity is invariably measured with respect to some "assumed" frame of reference. In electrical network theory, the frame of reference takes the form of an assumed reference direction of the current i and an assumed reference polarity of the voltage v. A thorough understanding of the concept of reference current direction and reference voltage polarity is absolutely essential in the study of nonlinear network theory. It is a fact that a large percentage of the mistakes committed by students of network theory can be traced to either the students' underestimation of the full significance of reference current directions and voltage polarities or the students' failure to maintain a consistent set of references.

Perhaps the simplest way to introduce the concept of assumed reference direction and polarity is through the following experiment. Suppose we are given a black box with a pair of terminals *a-b* and a wire *c-d* coming out of the box as shown in Fig. 1-2*a*. Suppose we are required to measure the voltage between terminals *a-b* and the current in the wire *c-d*.

Let us consider first measuring the voltage by connecting terminals *a-b* to the vertical input terminals of an oscilloscope.

---

[1] It can be justified on physical grounds that the simultaneity postulate is generally valid if the largest physical dimension of the device is much smaller than the *wavelength* of the highest anticipated frequency of operation.

Since one of the two vertical input terminals of any oscilloscope is marked with a positive sign while the other is marked with a negative sign, the question that immediately arises is which of the two terminals of the black box should we connect to the positive terminal of the oscilloscope in order to obtain the desired information. The answer is that it does not matter. In order to see this, suppose we *arbitrarily assume* terminal b to be connected to the positive terminal as shown in Fig. 1-2b. The assumption that terminal b is the positive terminal does not mean that the potential at b is higher than the potential at a. It does mean, however, that if at any time $t = t_1$, $v(t_1) > 0$, then the potential at b is higher than the potential at a. On the other hand, if $v(t_1) < 0$, then the potential at b at $t = t_1$ is actually lower than the potential at a. For example, if the voltage $v(t)$ displayed on the oscilloscope is given by

$$v(t) = 10 \sin \pi t \qquad \text{volts}$$

then terminal b is at a higher potential than terminal a during the time interval $0 < t < 1$ sec. But during the time interval $1 < t < 2$ sec, terminal b is actually at a lower potential than terminal a.

Let us now consider what happens when we assume terminal a instead of terminal b to be the positive terminal, as shown in Fig. 1-2c. Since the connection in Fig. 1-2c is opposite to the connection in Fig. 1-2b, it is clear that the voltage $v(t)$ displayed on the oscilloscope is now given by

$$v(t) = -10 \sin \pi t \qquad \text{volts}$$

Since terminal a is now the assumed positive terminal, and since $v(t) < 0$ for $0 < t < 1$ sec, this means that during this time interval, terminal a is at a lower potential than terminal b. Similarly, we found that during the time interval $1 < t < 2$, terminal b is actually at a lower potential than terminal a.

In either case we found the final answers to be identical. We can, therefore, conclude that in order to specify the voltage between any pair of terminals unambiguously, we may arbitrarily assume any one of the two possible terminals to be the positive terminal.

By analogy, we can conclude that in order to specify the current in any wire unambiguously, we may arbitrarily assume any one of the two possible directions to be the positive direction. The actual direction in which the current $i(t)$ is flowing at any time $t = t_1$

will be in the assumed positive direction if $i(t_1) > 0$, and opposite to the assumed direction if $i(t_1) < 0$.

Let us consider next a two-terminal black box N and assume a reference direction for the terminal current i and a reference polarity for the terminal voltage v. Since the references for both i and v are arbitrary, there are four distinct sets of combinations of references. There is no reason to prefer any one combination over the others. However, in practice, it is usually convenient to choose the combination so that *positive* power

$$p(t) = v(t)i(t) > 0$$

represents power *entering* the black box. From basic electromagnetic principles, it can be shown that this condition is satisfied whenever the current is chosen to enter the assumed positive terminal of the black box. From the simultaneity postulate, the same current must leave the negative terminal. This means that the allowable reference combination must be either of the form shown in Fig. 1-3*a* or *b*.

In either case, observe that the current arrow either enters the positive terminal or leaves the negative terminal.

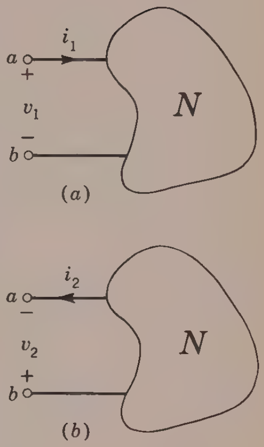


Fig. 1-3. Two possible sets of assumed reference direction and polarity for ensuring that positive power means power entering the black box.

1-4 INDEPENDENT SOURCES

Electrical energy must be supplied in order to move the charges which constitute the current i. Since energy can be neither created nor destroyed, it must be transformed from some other forms of energy. For example, a battery transforms chemical energy into electrical energy, a generator transforms mechanical energy into electrical energy. For convenience, we often refer to these energy-transforming devices, such as batteries or generators, as *sources of electrical energy* or simply *sources*. However, this statement should not be interpreted as implying that sources can create energy.

One of the earliest devices which serves as a source of electrical energy is the galvanic voltaic cell. Many other devices have been invented to function as sources of electrical energy, and, no doubt, many more will be invented in the future. Perhaps the simplest source of electrical energy today is the battery, which is capable of delivering a limited range of direct current to an external load connected with it, while maintaining an approximately *constant voltage* across its terminals. A less common source of electrical energy (but one gaining in popularity) is the solar cell, which develops a limited range of voltage drop across an external load connected with it, while maintaining an approximately *constant current* in the load. Observe that in the case of the battery, the output voltage is independent of the current drawn by the load (provided the current is not large). In contrast with this, it is the output current of the solar cell that is independent of the voltage drop across the load (provided the voltage drop is not large). It seems reasonable, therefore, to distinguish the above two types of electrical sources by calling the former a *voltage source* and the latter a *current source.*

Observe that in the above discussion, we have assumed that the current in the battery and the voltage drop across the solar cell are not large. This assumption is necessary because when a large current is drawn from the battery, its output voltage decreases and is no longer independent of the current. Similarly, a large voltage drop across the solar cell results in a decrease in the output current. The above phenomenon is a well-known experience; for example, the light dims whenever an appliance such as an air conditioner (which draws a large current) is turned on. In fact, no physical voltage source exists which is capable of developing a voltage that is entirely independent of its terminal current. Neither does there exist any physical current source which is capable of delivering a current that is entirely independent of its terminal voltage. While no such physical sources really exist, it is, nevertheless, extremely convenient to postulate the existence of the above sources as "ideal" sources. In other words, we are trying to "model" a physical source by an "ideal source" so that a network containing such sources can be conveniently analyzed. Observe that the above concept of modeling is analogous to that of the physicist who tries to represent the motion of a physical object by the motion of a "point" representing the center of gravity of the object. The concept of making a model to represent a physical system is so basic that we shall have many more occasions to use it in the future. With the above clarification, let us now render the concepts of independent sources more precise by the following discussions.

Independent voltage source An independent voltage source is a two-terminal device whose terminal voltage v is always equal to some given function of time $v_s(t)$, regardless of the value of the current flowing through its terminals; for example, $v_s(t) = 2 \sin t$. In particular, $v_s(t)$ may be a constant function such as $v_s(t) = E$, in which case, by analogy with direct current (dc) sources, we shall call the voltage source a dc voltage source. We shall use the symbols shown in Fig. 1-4 to denote an independent voltage source. Observe that a dc voltage source is denoted by the standard battery symbol in order to conform to popular usage.

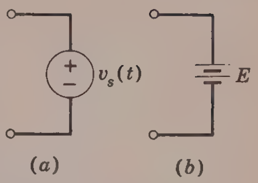


Fig. 1-4. Symbols for an independent voltage source.

Independent current source An independent current source is a two-terminal device whose terminal current i is always equal to some given function of time $i_s(t)$, regardless of the value of the voltage across its terminals. In particular, $i_s(t)$ may be a constant function such as $i_s(t) = I$, in which case we shall call the current source a dc current source. We shall use the symbols shown in Fig. 1-5 to denote an independent current source. Observe that a dc current source is denoted by the same symbol with the exception that $i_s(t)$ is replaced by a constant, I, independent of time.

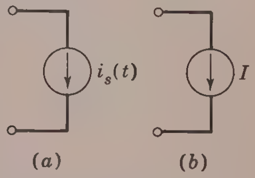


Fig. 1-5. Symbols for an independent current source.

Exercise: It is sometimes convenient to define an *independent flux-linkage source* and an *independent charge source* for the remaining two variables φ and q. State an analogous definition for each.

1-5 CHARACTERIZATION OF A TWO-TERMINAL BLACK BOX

Among the many physical devices of various complexities we shall be concerned in this chapter only with those which possess two *accessible* electrical terminals. Actually, the device may contain more than two terminals but only two of these are accessible to the external world in the sense that the device may be excited only through these terminals. For our purpose, it is convenient to imagine that the device is enclosed in a box and that the two accessible terminals are brought out by two connecting wires as shown in Fig. 1-6*a*. We shall call the resulting system a *two-terminal black box* and shall denote it by the symbol shown in Fig. 1-6*b*. It is important to emphasize that the *content* of the box may be as simple as a light bulb or as complicated as an arbitrary interconnection of other black boxes as shown in Fig. 1-6*c*.

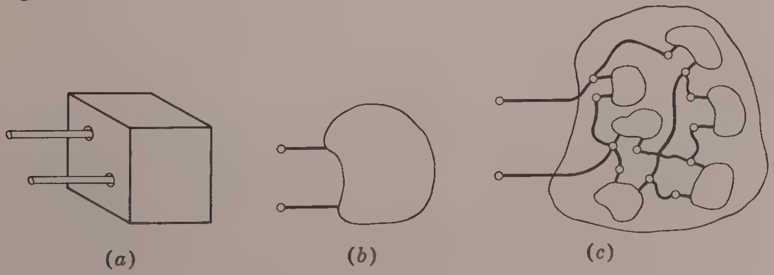
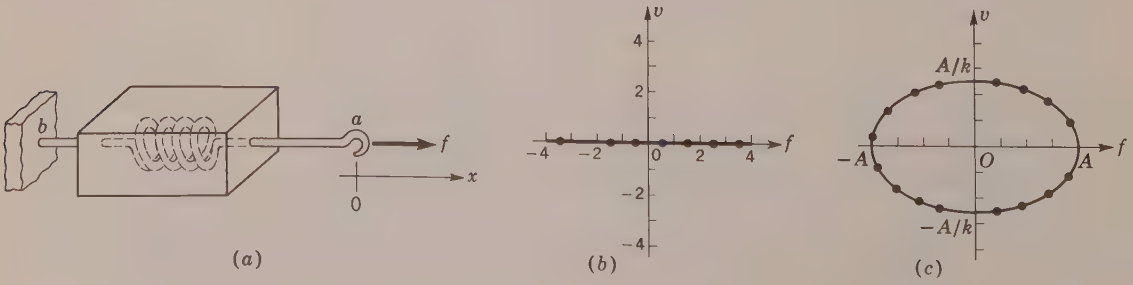


Fig. 1-6. Symbolic representation of a two-terminal black box.

The choice of the term “black box” is quite appropriate here because the box is really black inside in the sense that we cannot see its contents. As a matter of fact, unless we open the box and peep inside, there is no way of determining its contents. However, as engineers, we are not so much interested in the contents of the box as in knowing what the black box can do and how it behaves externally when it is connected with other black boxes into a network. In other words, we are primarily interested in predicting the external behavior of the black box without having to perform any tedious experiment. Our first step toward such an analytical approach is to “characterize” the black box. The concepts involved in characterizing a black box are so important that we pause here to consider a simple but illustrative analogy.

1-5-1 A MECHANICAL BLACK-BOX ANALOGY

Suppose we are given the mechanical black box containing a “spring” as shown in Fig. 1-7*a*. Suppose we did not know the contents of this black box and were asked to predict the behavior of the external terminals when an arbitrary force $f(t)$ is applied to terminal a of the black box with terminal b fixed against the wall. The mechanical variables of interest here are the displacement x (displacement to the right of the initial position 0 is assumed positive), the velocity v (of terminal a), and the force f (positive for tension and negative for compression). Clearly, the only way we can hope to characterize this black box (other than opening the box) is to start performing some experiments. Suppose we begin by applying a constant force $f = A$ and measure the corresponding velocity of terminal a. This would give us a point in the velocity-vs.-force plane (f-v plane). By repeating the above experiment with several values of force f, we obtain the data shown in Fig. 1-7*b*. We might be tempted to draw a smooth curve through



(*a*) (*b*) (*c*)

Fig. 1-7. An example illustrating the characterization of a mechanical black box. The data points in the v-vs.-f plane were found to lie on the horizontal axis in (*b*) and on the ellipse in (*c*).

these data points (which in this case happen to be the f axis) and claim to have characterized the black box in the sense that given any constant force f, we can analytically predict the associated velocity. However, a little thought will show that we have not really characterized the black box yet, for if, instead of applying a constant force we apply a slowly varying sinusoidal force, $f(t) = A \sin t$. The above characteristics would predict that $v(t) = 0$. This is of course contrary to what we expect to observe experimentally; namely, $v(t) = (A/k) \cos t$ where k is the "spring constant." We might hope that this inconsistency can be resolved by plotting all points (f,v) satisfying the above equations and obtaining an ellipse as shown in Fig. 1-7*c*. Observe, however, that the length of both axes of the ellipse depends on the amplitude A of the sinusoidal force, and for each value of A we would obtain a corresponding ellipse, so that eventually the entire f-v plane would be filled up with data points. Moreover, even if we can draw an infinite set of ellipses, we would be able to predict the velocity only if $f(t)$ is sinusoidal. Using these ellipses to predict v due to nonsinusoidal $f(t)$ would again yield erroneous answers. Reluctantly, we must admit that our efforts so far have been in vain and that just about the only useful information we obtained from the above experiment is that the black box cannot be characterized by a curve in the f-v plane.

Let us try another set of variables, say the force f and the displacement x, and repeat the experiments. As before, we begin by applying a constant force $f = A$ and measure the corresponding displacement x. Repeating this for various values of f, we obtain the data points shown in Fig. 1-8*a*. If we draw a smooth curve through these points, we obtain a single relationship

$$x = T(f)$$

Before we try to draw any conclusion, however, our previous experience suggests that we repeat the experiment with time-varying forces to see whether the above relationship still holds. Carrying out the proposed experiment with several low-frequency sinusoidal waveforms as before, we find that at any time $t = t_0$, the data point $[f(t_0),x(t_0)]$ always falls on the same curve $x = T(f)$. This is very encouraging, but to be sure, we must try some other nonsinusoidal waveforms for $f(t)$. Again, we find that, provided $f(t)$ does not change very rapidly,[1] the data point at any time also agrees with the curve in Fig. 1-8*a*. Hence, we can now draw the following conclusion: *For any f(t) which does not change rapidly,*

---

[1] This condition is actually equivalent to the statement that the frequency of the sinusoidal waveform is not very high. This will become obvious after the reader studies signal analysis.

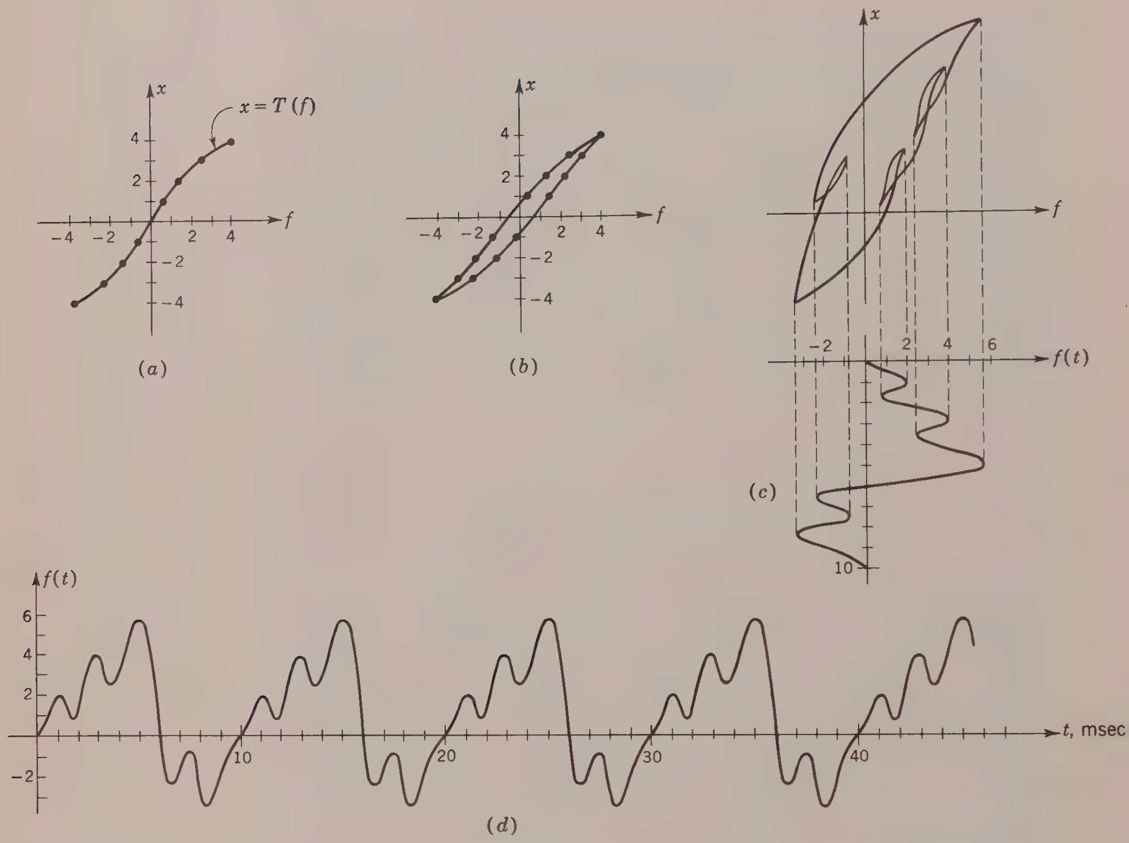


Fig. 1-8. As the frequency of the forcing function $f(t)$ increases, the x-vs.-f characteristic of the mechanical black box changes from a monotonic curve to a hysteresis loop.

the black box can be characterized by the displacement-vs.-force (f-x) curve shown in Fig. 1-8a.

After experiencing the length of time needed to carry out the above experiments, we can now begin to appreciate the utility of such a conclusion; namely, the characterization of the black box permits an analytical solution and thereby eliminates the need to carry out any further experiments.

Observe that our conclusion is based on the assumption that $f(t)$ does not change rapidly. Let us now repeat our experiment with higher-frequency sinusoidal waveforms, as well as with nonsinusoidal waveforms which change rapidly. The experiment shows that as we increase the frequency of the sinusoidal force $f(t)$, the data points begin to deviate (rather slowly at first) from the predicted curve $x = T(f)$. As we increase the frequency further, the data points begin to form a closed loop as shown in Fig. 1-8b, and the area enclosed by the loop tends to increase with frequency. Similarly, we find that if we apply a nonsinusoidal force which changes rapidly with time, the deviation from the curve in Fig. 1-8*a* is even worse. For example, Fig. 1-8*c* shows the *f-x* curve corresponding to the high-frequency nonsinusoidal waveform shown in Fig. 1-8*d*. The above experimental result shows that our earlier assumption, that $f(t)$ should not change very rapidly, is indeed necessary. In order to emphasize this restriction, it is a common practice to call the relationship obtained in Fig. 1-8*a* a *static* characteristic curve in contrast to the *dynamic* characteristic curve which corresponds to measurements at higher frequencies. Since the deviation of the measured characteristic curve from the static characteristic increases slowly with frequency, rather than abruptly, it is impossible to pick a definite frequency above which the static characteristic does not hold. Neither is it possible to find a single dynamic characteristic curve which would hold for all high frequencies. Hence, a certain amount of engineering judgment is involved in deciding whether a certain static characteristic curve can be used satisfactorily to solve a given problem. It is encouraging, however, to know that a large percentage of practical networks can indeed be analyzed by using only static characteristics. Moreover, even in cases when the static characteristic fails to give satisfactory solutions, we shall show in the future that we can often patch up the error by including "parasitic elements," namely, elements which are undesirable but which are invariably present in the black box in minute quantities. For the above example, the parasitic element consists of the *mass* associated with the spring. At low frequencies, the mass, being quite small, has relatively no effect on the measured *f-x* characteristic. However, as the frequency of the external force $f(t)$ increases, the acceleration of the spring increases, and the inertia force due to the mass becomes appreciable and, in fact, increases as acceleration increases. The deviation of the dynamic characteristic in Fig. 1-8*b* and 1-8*c* from the static characteristic in Fig. 1-8*a* can therefore be attributed to the inertia mass of the spring.

1-5-2 STATIC CHARACTERISTICS OF A TWO-TERMINAL BLACK BOX

The above discussion clearly shows the significance of static characteristics of a black box. Since *all characteristics to be considered in this book are assumed to be static characteristics,* we shall henceforth delete the adjective "static."

Let us now return to the main theme of this section, namely, the characterization of a two-terminal black box. Clearly, the only way we can hope to achieve this is to perform some meaningful external measurements. The only quantities of interest to us are those which can be measured externally. For example, the terminal voltage v and the terminal current i are of primary interest because they can be readily measured. The charge q and the flux linkage φ are also of interest because they can be indirectly measured by *integrating* the measured current waveform $i(t)$ and the measured voltage waveform $v(t)$ in accordance with Eqs. (1-11) and (1-12), respectively. From these measurements, we shall then try to establish a relationship, if there is any, between each pair of *independent* variables. Since the members of the pair of variables i and q are related by Eq. (1-7), they are not independent. Similarly, the variables v and φ are related by Eq. (1-8) and are also not independent. The only remaining combinations consist, therefore, of a relationship between the following variables:

1. Relationship between v and i
2. Relationship between v and q
3. Relationship between i and φ
4. Relationship between q and φ

The last relationship does not occur frequently in practice and has little practical significance. Hence, we shall restrict our attention throughout this book to only the first three cases. These correspond, respectively, to three basic types of two-terminal network elements, namely, a *two-terminal resistor,* a *two-terminal capacitor,* and a *two-terminal inductor.*

Our next step is therefore to plot the data in the v-i, v-q, and i-φ planes, in order to see if the points in any one of these planes can be connected to form a curve. In general, this may not be possible. For example, suppose the two-terminal black box happens to be a capacitance of 1 F. Then from elementary physics, we know that the relationship between v and i is $i = 1(dv/dt)$. But suppose we did not know that the black box contains a capacitance and proceeded to plot the data in the v-i plane. Clearly, it is impossible to expect that a curve can be found which passes through all data points in the v-i plane; in fact, if we take enough measurements, the data points will eventually fill the entire v-i plane. This is easily seen if we apply a voltage source of the form $v(t) = A \sin t$; since $i = dv/dt$, we obtain $i(t) = A \cos t$. Hence at any time $t = t_0$, we obtain a point $(A \sin t_0, A \cos t_0)$ in the v-i plane. Observe next that corresponding to each value of A, the above points form a circle of radius A since $v^2 + i^2 = A^2$. If we vary the value of A from 0 to ∞, we would eventually fill up the v-i plane with data points, and it would be impossible to find a curve passing through these points. On the other hand, if we choose to plot the points in the v-q plane, then these points can be connected by a smooth curve, namely, the line $q = v$. Therefore if a curve can be found which passes through all possible data points in either the v-i, the v-q, or the i-φ plane, then the two-terminal element is completely characterized by that curve.

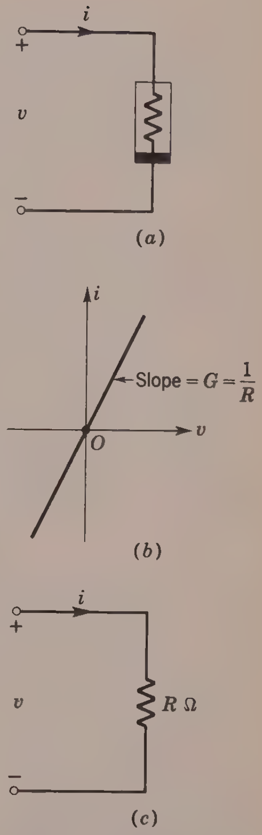


Fig. 1-9. Symbols for a two-terminal resistor.

1-6 TWO-TERMINAL RESISTORS

A two-terminal black box which can be characterized by a curve in the v-i plane is called a *two-terminal resistor* and will be denoted by the symbol shown in Fig. 1-9*a*. Observe that one edge of the symbol is darkened in order to distinguish between the two terminals. This is necessary because the v-i curve measured across the two terminals of a resistor is generally different from that measured across *the same* resistor but with the terminals interchanged (see Prob. 1-1).[1]

1-6-1 LINEAR RESISTORS

Among the infinite variety of v-i curves there is an important subclass which consists of *straight lines passing through the origin* as shown in Fig. 1-9*b*. Resistors of this subclass are called *linear resistors* and will be denoted by the standard symbol shown in Fig. 1-9*c*. Since the v-i curve of a linear resistor is a straight line *through the origin,* it can be described mathematically by $i = Gv$, or $v = Ri$. The constant G represents the slope of the line and is called the *conductance.* The constant R is defined as the reciprocal of G and is called the *resistance.* The practical unit of conductance is the *mho*. The practical unit of resistance is the *ohm* and will be denoted by Ω. A linear resistor is therefore completely characterized by one number, its conductance or its resistance. If the value of the resistance is positive, the linear resistor is said to be a *positive resistor.* Otherwise, it is said to be a *negative resistor.* If $R = 0$, the linear resistor is said to be a *short circuit.* If $R = \infty$, it is said to be an *open circuit.*

---

[1] In view of the nonsymmetrical nature of this symbol, we may avoid drawing voltage polarity and current direction signs beside the symbol *provided* we agree to assume that the darkened edge is the negative terminal and that the current enters the positive terminal. This convention will be followed in this book.

It is important to differentiate between the terms *resistor* and *resistance;* the former refers to a black box, but the latter refers to a property associated with the black box.[1]

Exercise 1: Explain why it is unnecessary to differentiate between the terminals of the symbol for a linear resistor.

Exercise 2: A certain v-i curve is described by an equation $v = 10i + 5$. Is this a linear resistor?

1-6-2 NONLINEAR RESISTORS

If a resistor is characterized by a v-i curve other than a straight line through the origin, it is called a *nonlinear resistor.* In this case, the resistor can no longer be described by a single number, and hence the entire v-i curve must be given. This may be specified either graphically by a curve or analytically by a mathematical relationship. For example, consider the set of practical two-terminal resistors listed in Table 1-1.[2] Since these components are all commercially available, they have been given names and symbols.[3] Each resistor in this table is characterized graphically by a typical v-i curve usually supplied by the manufacturer. In some cases, it may be possible to derive a mathematical relationship which closely approximates a certain v-i curve. For example, from physical principles one can show that the v-i curve of a vacuum diode can be represented approximately by a $3/2$ power law, namely,[4]

$$i = kv^{3/2} \tag{1-13}$$

where k is a constant which depends on the physical dimensions of the internal structure of the diode. Similarly, a semiconductor junction diode can be represented approximately by an exponential law, namely,[5]

$$i = I_0(e^{kv} - 1) \tag{1-14}$$

where I_0 and k are constants which depend on the physical parameters of the diode. One can also sometimes derive an equation which approximates a v-i curve by interpolation and approximation techniques (see Appendix A). For example, the varistor shown in Table 1-1 can be represented approximately by the equation

$$v = \alpha i^{\beta} \tag{1-15}$$

---

[1] This subtle difference is not universally recognized. In many books, the terms resistor and resistance are used synonymously. In this book, the term resistance refers only to a linear resistor.

[2] To conform with the IEEE standard letter symbols for semiconductor devices (*IEEE Trans. Electron Devices,* vol. Ed-11, no. 8, pp. 392–397), we have chosen the *uppercase letters* V and I in favor of the *lowercase letters* v and i as used in the context. Whenever applicable, we shall also follow the latest IEEE standards for graphic symbols.

[3] In view of its relatively recent origin, the name and symbol for the constant-current diode are not universally used. The same device is sometimes referred to as a *current-limiting diode, a currector, a field-effect diode,* etc. A further discrepancy may be found in that portion of the v-i curve for negative voltages. Depending on how the device is made, the v-i curve for $v < 0$ either approximates an open circuit (horizontal line), *as will be assumed throughout this book,* or a short circuit (vertical line). Fortunately, this discrepancy is usually not important because, as will be shown later, only the portion of the v-i curve in the first quadrant is actually of practical interest. However, in any case, if the v-i curve for $v < 0$ approximates a short circuit, it can always be transformed into the v-i curve shown in Table 1-1 by connecting a junction diode in series

where α and β are constants which can be determined numerically from the curve. In all cases, we must remember that any mathematical relationship is at best an approximation to the actual *v-i* curve. Moreover, most *v-i* curves cannot be represented by such simple expressions as those given above. Therefore, the most general and common method to specify element characteristics is to describe the curve in graphical form.

1-6-3 CLASSIFICATION OF *v-i* CURVES

In order to be able to use the nonlinear resistors effectively in a practical design, it is necessary to classify *v-i* curves into various categories. For example, the *v-i* curves of the first three resistors in Table 1-1 have one property in common; namely, for each pair of points (v_1,i_1) and (v_2,i_2) on the curve, we observe that whenever $v_1 > v_2$, then $i_1 > i_2$. Such elements are said to be *strictly monotonically increasing* resistors. An examination of the *v-i* curves of the zener diode and the constant-current diode shows that they are not strictly monotonically increasing because if we pick a pair of points with voltages $v_1 > v_2$ along the horizontal portions of the *v-i* curve, then $i(v_1) \not> i(v_2)$. However, these *v-i* curves have another common property; namely, $i(v_1) \geq i(v_2)$ for any $v_1 > v_2$. Such elements are said to be monotonically (but not strictly) increasing resistors. The *v-i* curves of the tunnel diode and the remaining resistors below it are not monotonically increasing because each *v-i* curve has a portion having *negative slopes* ($di/dv < 0$). Such elements are sometimes called *negative-resistance elements*. Another common characteristic of a negative-resistance element is that either the voltage is a multivalued function of current (more than one voltage corresponds to some given value of current) or the current is a multivalued function of voltage (more than one current corresponds to some given value of voltage). In the first case, the current is a single-valued function of the voltage (but not vice versa); that is,

$$i = i(v) \tag{1-16}$$

and is therefore called a *voltage-controlled resistor*. In the second case, it is the voltage which is a single-valued function of current; that is,

$$v = v(i) \tag{1-17}$$

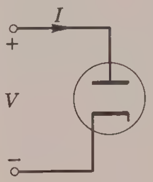
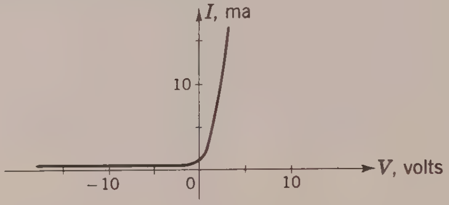
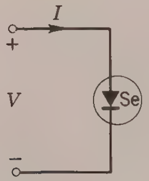
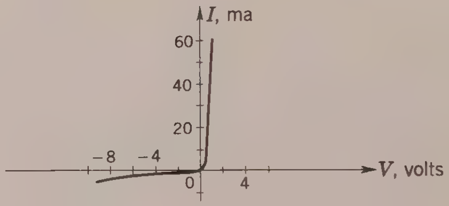
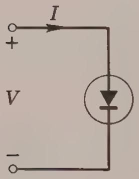
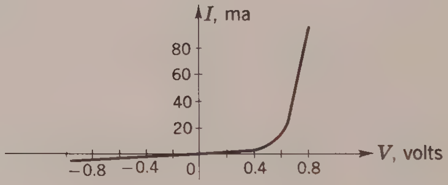
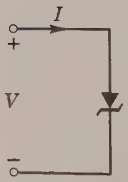
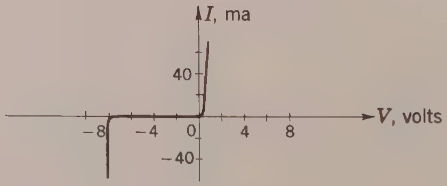
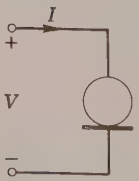
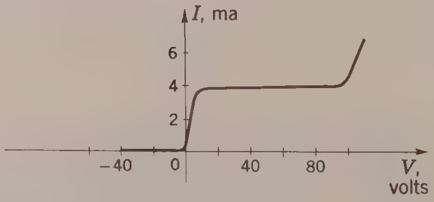
---

with the constant-current diode. For more information concerning this device, the reader is referred to J. M. Carroll, "Microelectronic Circuits and Applications," pp. 234 and 235, McGraw-Hill Book Company, New York, 1965; and J. M. Carroll, "Tunnel-Diode and Semiconductor Circuits," pp. 122–128, McGraw-Hill Book Company, New York, 1963.

[4] J. Langmuir, The Effect of Space Charge and Residual Gases on Thermionic Currents in High Vacuum, *Phys. Rev.*, vol. 2, pp. 450–486, 1913.

[5] J. F. Gibbons, "Semiconductor Electronics," McGraw-Hill Book Company, New York, 1966.

TABLE 1-1 *Practical two-terminal resistors.*

Name	Symbol	v-i characteristic curve
Vacuum diode		
Selenium diode		
Semiconductor (junction) diode		
Zener (avalanche, breakdown) diode		
Constant-current diode		

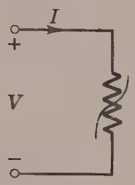
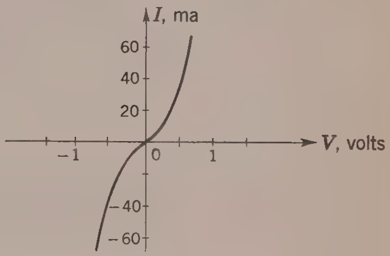
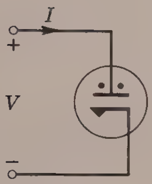
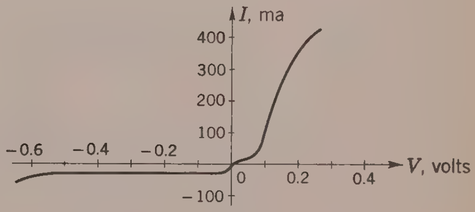
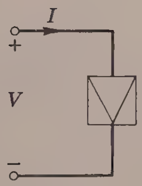
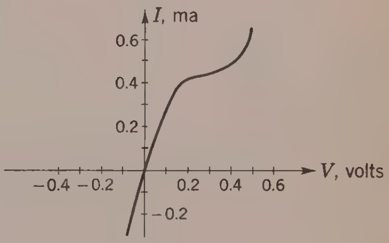
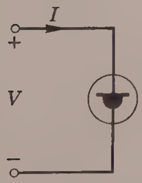
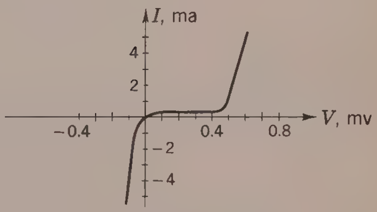
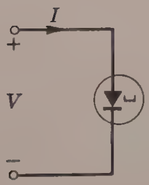
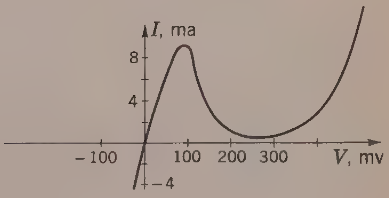
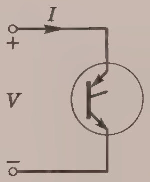
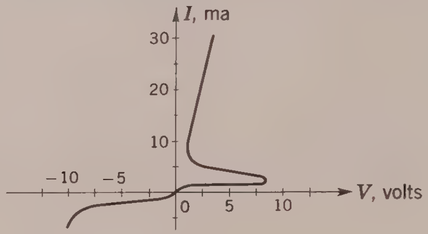
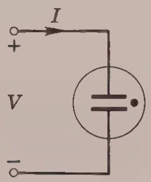
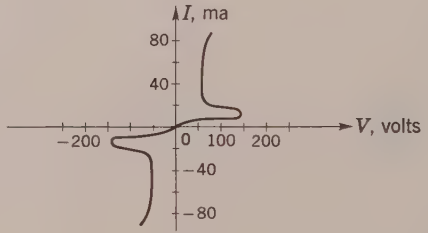
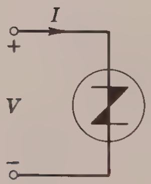
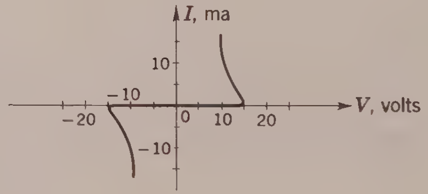
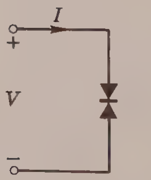
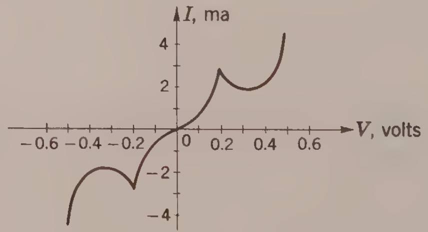
Name	Symbol	*v-i* characteristic curve
Varistor		
Solion liquid diode		
Tunnel resistor		
Back diode		
Tunnel diode		

TABLE 1-1 (*Continued*)

Name	Symbol	v-i characteristic curve
Four-layer diode		
Glow tube		
Trigger diode		
Superconducting tunnel junction		

and is therefore called a *current-controlled resistor.*[1] For example, the tunnel diode is a voltage-controlled resistor but the glow tube is a current-controlled resistor. Observe that a strictly monotonically increasing resistor is both voltage-controlled and current-controlled and can therefore be described either in the form of Eqs. (1-16) or (1-17).

---

[1] More generally, any curve in the x-y plane is said to be an x-controlled curve if it is a single-valued function of x and a y-controlled curve if it is a single-valued function of y.

Another important property shared by some v-i curves is their symmetry with respect to the origin. Such elements are called *bilateral resistors* because in this case the two terminals may be interchanged without affecting the v-i curve (see Prob. 1-1). For the resistors listed in Table 1-1, the varistor, the glow tube, the trigger diode, and the superconducting tunnel junction are the only bilateral resistors. The rest are nonbilateral.

Exercise 1: It is sometimes convenient to describe a voltage-controlled resistor by an equation of the form $i = G(v)v$ and a current-controlled resistor in the form $v = R(i)i$. Find the functions $G(v)$ and $R(i)$ in terms of $i(v)$ in Eq. (1-16) and $v(i)$ in Eq. (1-17). Give a geometrical interpretation of $G(v)$ and $R(i)$.

Exercise 2: Is a monotonically (but not strictly) increasing resistor both current-controlled and voltage-controlled? If not, under what condition is it voltage-controlled? When is it current-controlled?

Exercise 3: A resistor which is neither voltage-controlled nor current-controlled is said to be a *multivalued resistor.* Give an example of a multivalued resistor. Can you describe a multivalued resistor in the form of Eq. (1-16) or (1-17)? Explain why. (See Appendix A.)

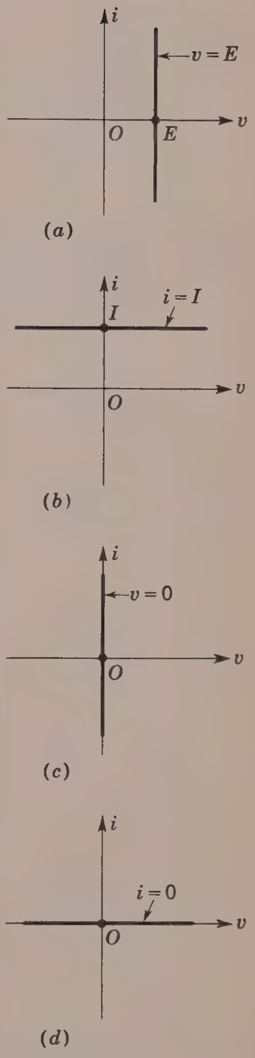
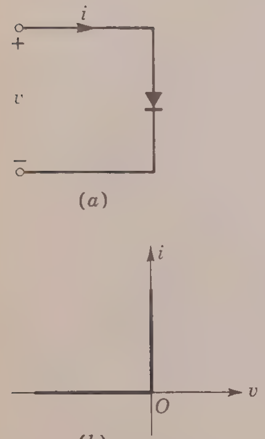


Fig. 1-10. The v-i curves of a dc-voltage source, a dc-current source, a short circuit, and an open circuit have one common property: they consist of either a vertical line or a horizontal line.

1-6-4 v-i CURVES OF DC SOURCES AND IDEAL DIODES

On many occasions we shall find it convenient to consider a dc voltage source and a dc current source as nonlinear resistors. This interpretation is valid because, by definition, a dc voltage source with terminal voltage E can be represented by a vertical line $v = E$ as shown in Fig. 1-10*a*. Similarly, a dc current source with terminal current I can be represented by a horizontal line $i = I$, as shown in Fig. 1-10*b*. In the special case where $E = 0$, the v-i curve of Fig. 1-10*a* becomes the $v = 0$ axis as shown in Fig. 1-10*c*. Since this coincides with the v-i curve of a short circuit, *a voltage source with zero terminal voltage is equivalent to a short circuit.* Similarly, when $I = 0$, the v-i curve of Fig. 1-10*b* becomes the $i = 0$ axis, as shown in Fig. 1-10*d*. Since this coincides with the v-i curve of an open circuit, *a current source with zero terminal current is equivalent to an open circuit.* Finally, a two-terminal resistor which does not exist in practice, but which is very useful conceptually, is the ideal diode whose symbol and v-i curve are shown in Fig. 1-11*a* and *b*, respectively. Analytically, an ideal diode is described by

$$\begin{aligned} i &= 0 \qquad \text{for all } v < 0 \\ v &= 0 \qquad \text{for all } i > 0 \\ p = vi &= 0 \qquad \text{for all } v \text{ and } i \end{aligned} \tag{1-18}$$



(a)

(b)

Fig. 1-11. The symbol and v-i curve of an ideal diode.

Observe that the last constraint is introduced to eliminate any point in the fourth quadrant from becoming a part of the v-i curve. It is also important to observe that an ideal diode becomes an open circuit for all $v < 0$ and a short circuit for all $i > 0$.

Before we leave this section, we wish to emphasize that the class of practical nonlinear resistors is not restricted to those listed in Table 1-1. In fact, when we reach Chap. 6, we shall be able to synthesize nonlinear resistors with almost any prescribed v-i curve of practical interest.

Exercise 1: Find the v-i curve of the ideal diode but with its terminals interchanged. Describe this curve analytically.

Exercise 2: A time-varying independent source may be represented by a family of v-i curves with the time t as a parameter. Sketch the v-i curves of a voltage source with terminal voltage $v_s(t) = 2t$ and a current source with terminal current $i_s(t) = 10 \sin \pi t$.

1-6-5 SOME PRACTICAL APPLICATIONS OF TWO-TERMINAL NONLINEAR RESISTORS

What are nonlinear resistors good for? How do we make use of their v-i curves to design practical electronic gadgets? Do certain types of v-i curves seem more appropriate for one application than another? These are some of the questions that will be answered in the latter part of this book, after we have built up enough theory to understand the basic principles involved in a practical design. However, to satisfy the impatient reader, we shall present in this section a qualitative description of some typical applications. Needless to say, this oversimplified treatment will become more quantitative and precise as the reader gains more ground in the subsequent chapters.

Rectification In many practical applications such as electroplating, the power supply must be restricted to a *single-polarity* voltage or current source. Since the most economical power source is 60-Hz sinusoidal voltage, it is desirable to transform this alternating voltage into a single-polarity voltage. This conversion process is called *rectification*, and any network that carries out the desired transformation is called a *rectifier*. The simplest rectifier consists of an ideal diode in series with a linear resistor, as shown in Fig. 1-12. When the input voltage $v_i(t)$ is positive, the diode becomes a short circuit and $v_o(t) = v_i(t)$. However, when the input voltage $v_i(t)$ is negative, the diode becomes an open circuit and $v_o(t) = 0$.

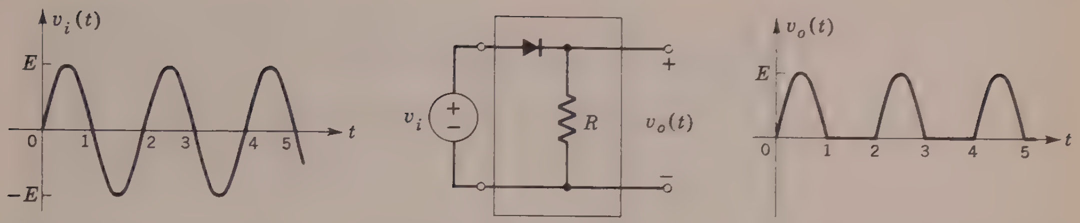


Fig. 1-12. An ideal rectifier converts a sinusoidal input voltage into a single-polarity output voltage.

The result is that the output voltage becomes zero during every other half cycle and is therefore a single-polarity voltage. Of course, this rectifier is an idealized circuit since it uses an ideal diode which does not exist in practice. However, an examination of Table 1-1 suggests that a practical rectifier may be constructed by replacing the ideal diode in Fig. 1-12 by a vacuum diode, a selenium diode, or a semiconductor junction diode.

The above procedure for arriving at a practical design by deriving first an idealized network (which is usually much easier to come by) and then approximating it by a practical circuit is a universal principle of creative design. This principle is based on the intuition that if two networks differ from each other only slightly (e.g., the *v-i* curves of corresponding resistors differ only slightly), then the corresponding voltage and current waveforms of the two networks must also differ only slightly. Mathematically, this is analogous to the variation of a *continuous* function; namely, a small variation in the value of the independent variable produces a correspondingly small variation in the value of the dependent variable. Because of its practical importance, we shall call the above assumption the small-variation postulate.

Frequency multiplication Another very common application of nonlinear resistors is to convert a low-frequency signal into a high-frequency signal. The ability to do this is instrumental in virtually all communication systems ranging from the simplest walkie-talkie to the most complex telemetry systems between artificial communication satellites. Amazingly, the principle for obtaining frequency multiplication is based on a simple observation from high school trigonometry; namely, the *n*th power of a sine or cosine function contains higher-harmonic components. For example, $\sin^3 x = \frac{3}{4}\sin x - \frac{1}{4}\sin 3x$, $\cos^4 x = \frac{3}{8} + \frac{1}{2}\cos 2x + \frac{1}{8}\cos 4x$, etc. Hence, if the *v-i* curve of a nonlinear resistor is described by a polynomial

$$i = a_0 + a_1 v + a_2 v^2 + a_3 v^3 + \cdots + a_n v^n \tag{1-19}$$

then upon applying a voltage signal $v = A \sin \omega t$, we obtain, with the help of various standard trigonometric identities, the expression

$$\begin{aligned} i(t) &= a_0 + a_1(A \sin \omega t) + a_2(A \sin \omega t)^2 + \\ &\qquad a_3(A \sin \omega t)^3 + \cdots + a_n(A \sin \omega t)^n \\ &= b_0 + b_1 \sin \omega t + b_2 \sin 2\omega t + b_3 \sin 3\omega t + \cdots + b_n \sin n\omega t \\ &\qquad + c_1 \cos \omega t + c_2 \cos 2\omega t + c_3 \cos 3\omega t + \cdots + c_n \cos n\omega t \end{aligned} \tag{1-20}$$

Observe that although the voltage consists of a sinusoidal signal of angular frequency ω, the resulting current contains a constant term b_0, a component of the same frequency ω, and a number of higher-harmonic components $2\omega, 3\omega, \ldots, n\omega$. In practice, any of these harmonic components can be extracted by interposing a network known as a *filter* which essentially suppresses all other components except the desired one.[1] In fact, we could even avoid the use of filters if we could obtain nonlinear resistors with suitable v-i curves. For example, to generate a third-harmonic signal, we apply a well-known trigonometric identity

$$\cos 3x = 4 \cos^3 x - 3 \cos x \tag{1-21}$$

to obtain the desired v-i curve,

$$i = 4v^3 - 3v \tag{1-22}$$

Hence, if $v = \cos \omega t$, then

$$i(t) = 4 \cos^3 \omega t - 3 \cos \omega t = \cos 3\omega t \tag{1-23}$$

which is the desired third harmonic. The next step then consists of finding a practical nonlinear resistor with a v-i curve which approximates Eq. (1-22). Unfortunately, no commercially available resistor is close enough even as an approximation. Hence, it would be necessary to *synthesize* this v-i curve using commercially available resistors as building blocks. The principles and techniques for synthesizing arbitrary v-i curves will be given in Chap. 8.

Exercise 1: Find the values of the coefficients $b_0, b_1, b_2, \ldots, b_n$ and $c_1, c_2, \ldots, c_n$ in Eq. (1-20) in terms of the constant A and the coefficients $a_0, a_1, a_2, \ldots, a_n$, where $n = 5$.

---

[1] The design of filters is a very well-developed subject and is usually given in a senior-level course called network synthesis.

Exercise 2: Using Eq. (1-21), find the desired v-i curve for converting a 100-volt, 60-Hz sinusoidal voltage into a 10-amp, 180-Hz sinusoidal current.

Exercise 3: Verify the trigonometric identity $\cos^5 x = \frac{5}{8} \cos x + \frac{5}{16} \cos 3x + \frac{1}{16} \cos 5x$ and find the desired v-i curve for converting a 1-ma, 1-kHz sinusoidal current waveform into a 1-volt, 5-kHz sinusoidal voltage waveform.

Frequency mixing Given two sinusoidal waveforms with *commensurate* angular frequencies ω_1 and ω_2 (that is, the ratio ω_1/ω_2 is a rational number), we are frequently interested in generating a new sinusoidal waveform with a frequency given by $(m\omega_1 \pm n\omega_2)$, where m and n are any integers, including zero. Each new frequency corresponding to a given combination (m,n) is called a *beat frequency* and will be denoted by $\omega_{(m,n)}$. One of the most common requirements in signal processing (e.g., a radio receiver or an electronic organ) is the generation of appropriate beat frequencies.[1] We shall now demonstrate that in order to generate beat frequencies, it is necessary to perform a nonlinear operation. Again, the basis for doing this is given by the well-known trigonometric identities:

$$\sin x \sin y = \tfrac{1}{2} [\cos (x - y) - \cos (x + y)] \tag{1-24}$$

and

$$\sin x \cos y = \tfrac{1}{2} [\sin (x + y) + \sin (x - y)] \tag{1-25}$$

To demonstrate how we generate beat frequencies, consider applying two voltage sources $v_1 = A \sin \omega_1 t$ and $v_2 = B \sin \omega_2 t$ in series with a nonlinear resistor with a v-i curve given by $i = v^3$. The current $i(t)$ is given by

$$\begin{aligned} i(t) &= (A \sin \omega_1 t + B \sin \omega_2 t)^3 \\ &= A^3 \sin^3 \omega_1 t + 3A^2 B \sin^2 \omega_1 t \sin \omega_2 t \\ &\qquad + 3AB^2 \sin \omega_1 t \sin^2 \omega_2 t + B^3 \sin^3 \omega_2 t \end{aligned}$$

If we now apply Eq. (1-25) and a number of standard trigonometric identities, we obtain, upon simplification, the expression

$$\begin{aligned} i(t) &= (a_1 \sin \omega_1 t + b_1 \sin \omega_2 t) + (a_2 \sin 3\omega_1 t + b_2 \sin 3\omega_2 t) \\ &\quad + [a_3 \sin (\omega_2 - 2\omega_1)t + b_3 \sin (\omega_2 + 2\omega_1)t] \\ &\quad + [a_4 \sin (\omega_1 - 2\omega_2)t + b_4 \sin (\omega_1 + 2\omega_2)t] \end{aligned} \tag{1-26}$$

---

[1] The beat frequency is also called a *sideband frequency* and the collection of all beat frequencies is usually called *sidebands.* The definitions of beat frequency and sidebands are meaningful even if ω_1 and ω_2 are not commensurate with each other.

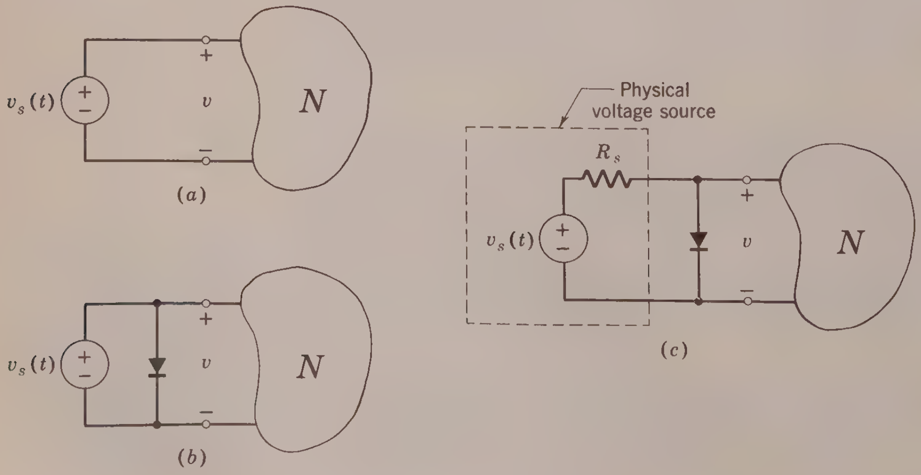
where the coefficients a_j, b_j are functions of A and B. Observe that in addition to sinusoidal terms having the same frequencies as the driving sources, the current $i(t)$ also contains the third-harmonic terms with frequencies $3\omega_1$ and $3\omega_2$ and the beat-frequency terms with frequencies $(\omega_2 \pm 2\omega_1)$ and $(\omega_1 \pm 2\omega_2)$. In the more general case where the v-i curve is described by a polynomial, we can expect, in general, sinusoidal terms with harmonic frequencies $m\omega_1$ and $n\omega_2$, as well as beat frequencies $m\omega_1 \pm n\omega_2$. In practice, any one of these beat frequencies may be extracted through a filter. This principle is widely used in telephone systems.

Exercise 1: Give an example of a pair of sinusoidal waveforms with incommensurate (i.e., not commensurate) frequencies. Is the sum of these two waveforms periodic?

Exercise 2: A speech synthesizer is an electronic system designed to simulate the human voice. An important component of this system is a mixer for generating as many beat frequencies as possible. Assuming that the v-i curve of the resistor is described by a polynomial, what must the general form of the polynomial be in order to generate beat-frequency terms with m and n equal to 0, ± 1, ± 2, and ± 3?

Limiting Any nonlinear resistor R with a v-i curve containing a (nearly) vertical segment can be used to *limit* the voltage across a two-terminal black box connected in parallel with R. For example, we can limit the terminal voltage across the black box N shown in Fig. 1-13a to nonpositive values by connecting an ideal diode across N as shown in Fig. 1-13b. This is because by definition, the voltage across an ideal diode is given by $v \leq 0$.

Fig. 1-13. The voltage across the black box N is constrained to nonpositive values by connecting an ideal diode in parallel with N.



(a)

(b)

(c)

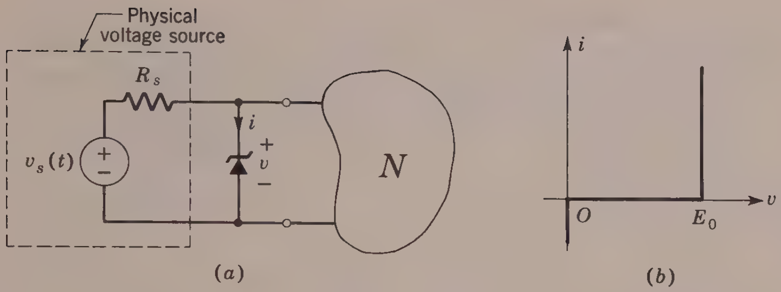
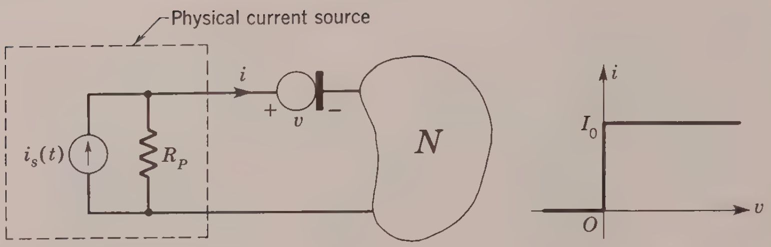


Fig. 1-14. The voltage across N is limited to a maximum value equal to the constant voltage E_0 of the zener diode.

A paradoxical situation arises when one questions what happens if a voltage source with a positive terminal voltage (say $v_s =$ 2 volts) is connected across the network shown in Fig. 1-13*b*. By definition, the terminal voltage of this voltage source must remain constant regardless of the external network connected across it. But also by definition, the voltage across an ideal diode cannot be positive. The basic problem here is that we are connecting two incompatible ideal elements in parallel, thereby rendering the definitions inconsistent. In other words, this paradox arises because of an overidealization. It is no different from many paradoxes of a similar nature, most notably among which is the paradox: "What happens if one connects a short circuit across a voltage source with a nonzero terminal voltage?" The best way to resolve this type of paradox is to exclude all such incompatible connections. But how can we forbid anyone from making an incompatible connection in practice? The answer is that there is no such thing as an incompatible connection *in practice* because there are no such things as an ideal voltage source and an ideal diode. Any physical voltage source has a small internal resistance R_s in series with it, as shown in Fig. 1-13*c*. Once we introduce R_s, the paradox disappears because whenever $v_s(t)$ becomes positive, the diode becomes a short circuit and the entire voltage appears across R_s. Hence, the voltage across N can never be positive.

The same principle can be applied to limit the voltage across N from exceeding a prescribed value E_0. For example, if we connect a zener diode with a constant voltage $E_z = E_0$ across N as shown in Fig. 1-14*a*, then from the v-i curve of the zener diode shown in Fig. 1-14*b* (observe that the reference polarity and directions are opposite to those shown in Table 1-1) it is clear that $0 \leq v \leq E_0$. This circuit is commonly used for overload protection. For example, in a typical application, the black box N consists of a sensitive instrument (such as a voltmeter) whose maximum permissible voltage is equal to E_0.

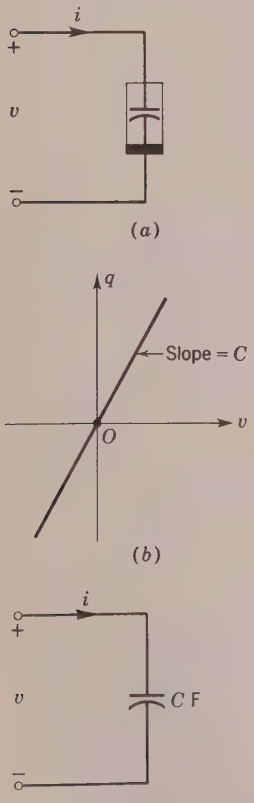


(a) (b)

Fig. 1-15. The current entering N is limited to a maximum value equal to the constant current I_0 of the constant-current diode.

By analogous reasoning, any resistor R with a v-i curve containing a (nearly) horizontal segment can be used to *limit* the current entering a black box connected in series with R. For example, if we connect a constant-current diode with constant current I_0 in series with N as shown in Fig. 1-15*a*, then from the v-i curve of the constant-current diode shown in Fig. 1-15*b* it is clear that $0 \leq i \leq I_0$ (the resistor R_p is introduced to avoid a similar paradox).

Fig. 1-16. Symbols for a two-terminal capacitor.



(a)

(b)

(c)

Exercise 1: The maximum permissible range of voltages of a hypersensitive instrument is given by $-10 \leq v \leq 5$. Design an overload protection circuit and specify the v-i curve of any nonlinear resistor used in the circuit.

Exercise 2: Explain what happens if $i_s(t) > I_0$ in the circuit shown in Fig. 1-15. Replace the constant-current diode with an appropriate nonlinear resistor so as to limit the terminal current entering N to $|i| < 20$ ma.

1-7 TWO-TERMINAL CAPACITORS

A two-terminal black box which can be characterized by a curve in the v-q plane is called a *two-terminal capacitor* and will be denoted by the symbol shown in Fig. 1-16*a*. Observe that one edge of this symbol is darkened for the same reason as it was for the resistor.

1-7-1 LINEAR CAPACITORS

An important subclass of capacitors can be characterized by a straight line through the origin of the v-q plane, as shown in Fig. 1-16*b*. This subclass is called *linear capacitors* and will be denoted by the conventional symbol shown in Fig. 1-16*c*. A linear capacitor can be described analytically by

$$q = Cv \tag{1-27}$$

where the constant C represents the slope of the straight line and is called the *capacitance* associated with the capacitor. The unit of capacitance is the *farad* and will be denoted by F. To find the current entering a linear capacitor, we substitute Eq. (1-27) for q in Eq. (1-7) and obtain

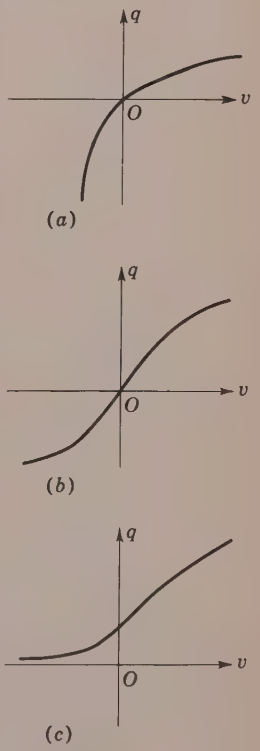
$$i(t) = C\frac{dv(t)}{dt} \tag{1-28}$$

A linear capacitor is therefore completely characterized by one number, namely, its capacitance. Again, we would differentiate between the terms *capacitor* and *capacitance.*

1-7-2 NONLINEAR CAPACITORS

If a capacitor is characterized by a v-q curve other than a straight line through the origin, it is called a *nonlinear capacitor.* In this case, the capacitor can no longer be described by a single number, and hence the entire v-q curve must be given. An example of a practical, nonlinear capacitor is the metal-oxide-semiconductor (MOS) capacitor whose v-q curve is shown in Fig. 1-17*a*. This nonlinear capacitor is used quite extensively in integrated circuits, where the conventional linear capacitor becomes impractical to fabricate. Although there are at present only a few practical nonlinear capacitors available commercially, it is expected that more will become available in the near future. In fact, as will be shown in Chap. 3, it is possible to synthesize a capacitor with any prescribed v-q curve with the help of a new network component called the *mutator.*

Fig. 1-17. The v-q curves of three typical nonlinear capacitors.



There are other reasons for studying nonlinear capacitors. One reason is that components of many physical and biological systems behave in a manner analogous to that of a nonlinear capacitor. Hence the study of such systems can often be achieved by constructing an electrical network model to simulate the behavior of these systems. A simple example is the displacement-vs.-force curve of the nonlinear spring shown in Fig. 1-7*a*. This mechanical element is usually modeled by a nonlinear capacitor with a similar v-q curve, as shown in Fig. 1-17*b*. Another example is given by the volume-vs.-pressure curve of the ventilatory part of the human respiratory system. This biological component can be modeled by an analogous v-q curve as shown in Fig. 1-17*c*.

We shall denote the v-q curve of a nonlinear capacitor by

$$q = q(v) \tag{1-29}$$

if it is voltage-controlled, and by

$$v = v(q) \tag{1-30}$$

if it is charge-controlled. For a voltage-controlled capacitor, the current entering the capacitor can be expressed in a form analogous to Eq. (1-28); thus

$$i(t) = \frac{dq(t)}{dt} = \frac{dq(v)}{dv}\frac{dv(t)}{dt}$$

or

$$i(t) = C(v(t))\frac{dv(t)}{dt} \tag{1-31}$$

where

$$C(v) \equiv \frac{dq(v)}{dv} \tag{1-32}$$

is called the *incremental capacitance* of the capacitor. Notice that the incremental capacitance is a function of the capacitor voltage and becomes a constant only in the case of a linear capacitor.

Exercise 1: A typical nonlinear capacitor is characterized by the v-q curve $q = kv^{3/2}$, where k is a physical parameter. (*a*) Find the incremental capacitance $C(v)$. (*b*) If the applied voltage is given by $v(t) = \frac{1}{4}\cos^2 t$, find the charge $q(t)$ and the current $i(t) \equiv dq(t)/dt$. (*c*) Calculate $i(t)$ by using Eq. (1-31).

Exercise 2: An abrupt-junction diode is a semiconductor p-n junction which behaves like a capacitor, provided the voltage across the junction is less than 0.5 volt. Its incremental capacitance is given by $C(v) = k(\phi - v)^{-1/n}$, where k, ϕ, and n are constants which depend upon the parameters of the device. (*a*) Plot the incremental capacitance on logarithmic paper for the range $-100 < v < 0.5$ volt. (Assume $k = 80 \times 10^{-12}$, $\phi = 0.5$, and $n = 2$.) (*b*) What are the maximum and the minimum values of the capacitance (in picofarads or 10^{-12} F) within this range of applied voltage? (*c*) Do you have sufficient information to recover the v-q curve? If not, what additional information do you need?

1-7-3 SOME PRACTICAL APPLICATIONS OF TWO-TERMINAL NONLINEAR CAPACITORS

What are nonlinear capacitors good for? Can they do useful things which nonlinear resistors cannot? The answer to the second question is obviously yes, for otherwise we would not be studying them. In addition to being able to do a number of things described earlier for resistors, a nonlinear capacitor can do better in certain cases. Although we do not yet have the background necessary to demonstrate this assertion, suffice it to say that both nonlinear resistors and capacitors are capable of generating higher harmonics. However, with an appropriate design, it is possible to extract more output power in any given harmonic component from a nonlinear capacitor than from a nonlinear resistor. This means that a nonlinear-capacitor-frequency multiplier has a higher efficiency than a nonlinear-resistor-frequency multiplier. In addition to this application, a few of the many other useful functions are briefly described as follows.

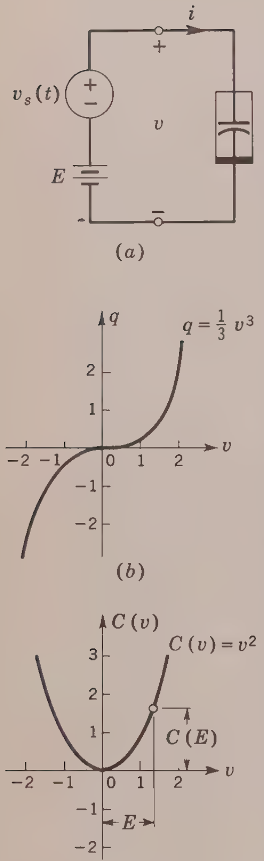
Frequency division In many practical systems, it is desirable to convert a given sinusoidal signal of frequency ω_1 into another sinusoidal signal of a lower frequency ω_2; namely, $\omega_2 = \omega_1/n$, where n is an integer. In this case, the lower-frequency output signal is said to be a *subharmonic* of the higher-frequency output signal. It can be shown that *a nonlinear resistor cannot generate subharmonics.* To demonstrate that a nonlinear capacitor can generate a subharmonic signal, consider a nonlinear capacitor whose incremental capacitance is given by

$$C(v) = \left[\frac{1 - \sqrt{1 - v^2}}{2(1 - v^2)}\right]^{1/2} \tag{1-33}$$

If we apply a voltage $v(t) = \sin \omega t$ across this capacitor, the current $i(t)$ can be calculated from Eqs. (1-31) and (1-33); thus

$$\begin{aligned} i(t) &= \left[\frac{1 - \sqrt{1 - \sin^2 \omega t}}{2(1 - \sin^2 \omega t)}\right]^{1/2} (\omega \cos \omega t) \\ &= \left(\frac{1 - \cos \omega t}{2 \cos^2 \omega t}\right)^{1/2} (\omega \cos \omega t) \\ &= \omega \sqrt{\frac{1 - \cos \omega t}{2}} = \omega \sin \frac{\omega}{2} t \end{aligned} \tag{1-34}$$

Hence, the output current is a sinusoid with frequency equal to half the original frequency. The phenomenon of subharmonic generation by a nonlinear capacitor has been utilized in many practical applications. One application consists of utilizing the two "distinct" frequencies as the two distinct states in designing a digital computer. Another interesting application consists of converting the high-frequency output of a laser beam into a lower-frequency signal.



(a)

(b)

(c)

Fig. 1-18. A nonlinear capacitor can be used as a tuning element by varying the dc voltage E across the capacitor.

Parametric amplifier Just as with nonlinear resistors, it is possible to generate beat-frequency components by applying two sinusoidal signals of frequencies ω_1 and ω_2 in series with a nonlinear capacitor. It can be shown that if one of the two signals (say $v_1 = A \sin \omega_1 t$) is very weak while the other signal (say $v_2 = B \sin \omega_2 t$) is very strong, it is possible to extract (with the help of filters) the signal with frequency ω_1 and at the same time greatly amplify its amplitude, say $v_o = 1{,}000\, A \sin \omega_1 t$. The result is that we have an amplifier. For reasons that we are not equipped to elaborate here, this amplifier is called a *parametric amplifier.* It is widely used in artificial satellites because it has some definite advantages over conventional amplifiers.

Electronic tuning Suppose we connect a voltage source $v_s(t)$ and a battery with terminal voltage E in series with a nonlinear capacitor as shown in Fig. 1-18*a*. For simplicity, let the v-q curve be given by $q = \frac{1}{3} v^3$ as shown in Fig. 1-18*b*. Then its incremental capacitance is given by $C(v) = v^2$, as shown in Fig. 1-18*c*. Now in many electronic systems, such as a radio receiver, the signal $v_s(t)$ is very small (say, a few millivolts) compared with the value of the dc voltage E. Hence, for most practical purposes, the incremental capacitance

$$C(v) = C(v_s(t) + E) \approx C(E) \tag{1-35}$$

can be considered to depend *only* on the value of E. In this case, Eq. (1-31) becomes

$$i(t) = C(E)\frac{dv}{dt} \tag{1-36}$$

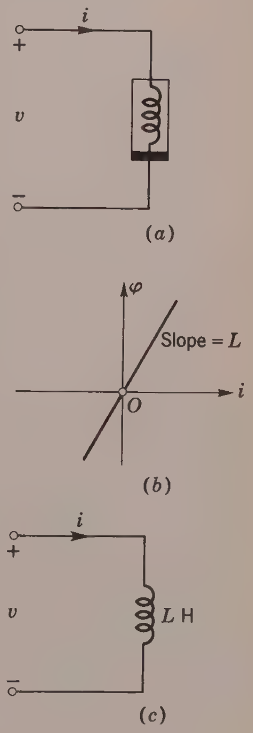
Since $C(E)$ is no longer a function of time, Eq. (1-36) is identical with Eq. (1-28) which describes a linear capacitor. The only difference is that we can change the value of the capacitance by simply changing the value of E. This observation is of great practical importance. One immediate application is in the area of *electronic tuning.* The conventional way to tune a radio receiver from one station to another is to turn a knob which moves the tuning dial. Any one who opens up the cover of a radio receiver would recognize that this tuning knob is used to rotate the plates of an air capacitor, thereby changing the value of its capacitance. In other words, the standard tuning process consists of adjusting the value of a capacitor *mechanically.* This operation can now be replaced by a nonlinear capacitor connected as shown in Fig. 1-18*a* where the tuning is accomplished by adjusting the voltage E. This method is clearly far superior to the use of bulky air capacitors. In fact, this technique of electronic tuning is fast becoming a standard method in electronic systems.

Exercise 1: Find the incremental capacitance $C(v)$ required to generate a 30-Hz subharmonic sinusoidal current waveform from an input voltage $v(t) = 100 \cos 120\pi t$. HINT: Make use of the trigonometric identity

$$\cos \frac{x}{2} = \sqrt{\frac{1 + \cos x}{2}}$$

Exercise 2: A common nonlinear capacitor used for electronic tuning is the *varactor diode*. It is characterized by a v-q curve $q(v) = -(3/2)C_0\phi_0(1 - v/\phi_0)^{2/3}$, where C_0 and ϕ_0 are constants which vary from device to device. When $v = 0$, the incremental capacitance was measured to be equal to 60 pF. (*a*) Derive the incremental capacitance $C(v)$. (*b*) If $\phi_0 = 0.35$, find the range of the input voltage required to tune the capacitance from 5 to 100 pF. To operate the varactor as a nonlinear capacitor, the voltage must not exceed 0.35 volt.

Fig. 1-19. Symbols for a two-terminal inductor.



(*a*) (*b*) (*c*)

1-8 TWO-TERMINAL INDUCTORS

A two-terminal black box which can be characterized by a curve in the i-φ plane is called a *two-terminal inductor* and will be denoted by the symbol shown in Fig. 1-19*a*. The darkened edge of this symbol has the same significance as before.

1-8-1 LINEAR INDUCTORS

An important subclass of inductors can be characterized by a straight line through the origin of the i-φ plane as shown in Fig. 1-19*b*. This subclass is called *linear inductors* and will be denoted by the conventional symbol shown in Fig. 1-19*c*. A linear inductor can be described analytically by

$$\varphi = Li \tag{1-37}$$

where the constant L represents the slope of the straight line and is called the *inductance* associated with the inductor. The unit of inductance is the *henry* and will be denoted by H. To find the voltage across a linear inductor, we substitute Eq. (1-37) for φ in Eq. (1-8) and obtain

$$v(t) = L\frac{di(t)}{dt} \tag{1-38}$$

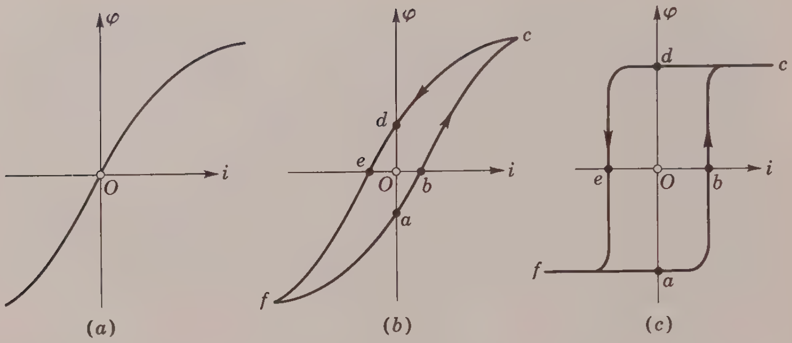


Fig. 1-20. The i-φ curve of three practical nonlinear inductors.

A linear inductor is therefore completely characterized by one number, namely, its inductance. Again, we would differentiate between the terms *inductor* and *inductance.*

1-8-2 NONLINEAR INDUCTORS

If an inductor is characterized by an i-φ curve other than a straight line through the origin, it is called a *nonlinear inductor.* In this case, the inductor can no longer be described by a single number, and hence the entire i-φ curve must be given. For example, Fig. 1-20*a* shows the i-φ curve of a typical nonlinear inductor.

Another common nonlinear inductor consists of a coil wound around an iron core. Its i-φ curve (obtained by applying a sinusoidal current excitation) is shown in Fig. 1-20*b*. This curve is a multivalued function of both i and φ and is commonly referred to as the *hysteresis loop.*[1] Observe that starting at point a with $i = 0$, the flux linkage φ increases with i along the path a-b-c. Upon reaching point c when φ attains its maximum value, the flux linkage φ does not retrace the original path. Instead, it decreases with the current i along the path c-d-e-f. Upon reaching point f when i attains its minimum value, the flux linkage φ returns to point a to complete the loop. The shape of the hysteresis loop depends on the type of material used for the core. For certain materials, the hysteresis loop is almost rectangular, as shown in Fig. 1-20*c*.

We shall denote the i-φ curve of a nonlinear inductor by

$$\varphi = \varphi(i) \tag{1-39}$$

if it is current-controlled, and by

$$i = i(\varphi) \tag{1-40}$$

---

[1] Actually, this hysteresis loop is a valid description only under the assumption that the current waveform is sinusoidal. For other periodic excitations, the hysteresis loop becomes much more complicated. A *complete* characterization of elements described by hysteresis loops is a very difficult and still unsolved problem.

if it is flux-controlled. In the case of a current-controlled inductor, the voltage across the inductor can be expressed in a form analogous to Eq. (1-38); thus

$$v(t) = \frac{d\varphi(t)}{dt} = \frac{d\varphi(i)}{di}\,\frac{di(t)}{dt}$$

or

$$v(t) = L(i(t))\,\frac{di(t)}{dt} \tag{1-41}$$

where

$$L(i) \equiv \frac{d\varphi(i)}{di} \tag{1-42}$$

is called the *incremental inductance* of the inductor. Notice that for a linear inductor, the incremental inductance coincides with the inductance, as it should.

Exercise 1: The i-φ curve of a certain nonlinear inductor can be represented approximately by the cubic equation $\varphi = i^3$. If the inductor is connected across a current source with terminal current $i_s(t) = \sin t$, find and sketch the incremental inductance $L(i)$ and the inductor voltage $v(t)$.

Exercise 2: An inductor is said to be the "dual" of a capacitor, and vice versa, because there exists a one-to-one correspondence between the two elements. Exhibit a list of corresponding quantities.

1-8-3 SOME PRACTICAL APPLICATIONS OF TWO-TERMINAL NONLINEAR INDUCTORS

What are nonlinear inductors good for? Where are they used in practice? To answer these questions would again require more background than we have at present. However, it is instructive to describe a few simple applications.

Frequency conversion Just as is true of capacitors, a nonlinear inductor is capable of generating both *harmonics* and *subharmonics* of a given sinusoidal signal. It can be shown to have the same efficiency as does a nonlinear capacitor. This property is widely used in telephone systems.

Memory and storage Consider the rectangular hysteresis curve shown in Fig. 1-20*c*. Observe that when $i = 0$, φ may assume either one of two distinct values (point *a* or point *d*) depending on the previous history of the excitation current. These two distinct states can be used to represent the two states (0 and 1) in a digital computer. When many of these elements are combined properly, the result is a "memory" or "storage" device to store present information for future use. While there are many other candidates, this memory device has some significant advantages. One is that in both states $i = 0$, and hence no power is being consumed. Since hundreds and thousands of these elements are used in a practical computer, the saving in power cost is enormous.

1-9 ENERGY AND POWER

The energy flow into a two-terminal black box during any time interval (t_0,t_1) is by definition the time integral of power from t_0 to t_1; namely,

$$w(t_0,t_1) = \int_{t_0}^{t_1} v(t)i(t)\,dt \tag{1-43}$$

Since $w(t_0,t_1)$ is a *relative* quantity depending on the time interval (t_0,t_1), it is convenient for us to define another related but *absolute* quantity by letting t_0 equal zero and t_1 approach infinity, and then take the average of the energy flow over the entire infinite time interval; namely,

$$P_{\text{av}} \equiv \lim_{t_1\to\infty} \frac{w(0,t_1)}{t_1} \tag{1-44}$$

Since the quantity P_{av} has the dimension of energy per second, it is called the average power. Substituting Eq. (1-43) into Eq. (1-44), we obtain the explicit expression

$$P_{\text{av}} = \lim_{t_1\to\infty} \frac{1}{t_1} \int_0^{t_1} v(t)i(t)\,dt \tag{1-45}$$

To illustrate the use of this formula, let us calculate the average power entering a 4-Ω linear resistor due to an applied voltage $v(t) = 2 \sin \pi t$; thus

$$P_{\text{av}} = \lim_{t_1\to\infty} \frac{1}{t_1} \int_0^{t_1} (2 \sin \pi t)\left(\frac{2 \sin \pi t}{4}\right) dt \tag{1-46}$$

$$P_{\text{av}} = \lim_{t_1\to\infty} \left(\tfrac{1}{2} - \frac{\sin 2\pi t_1}{4\pi t_1}\right) = \tfrac{1}{2}$$

In the case where the voltage $v(t)$ and current $i(t)$ are periodic functions with commensurate periods T_v and T_i, respectively, the power $p(t) = v(t)i(t)$ will also be periodic. However, the period of $p(t)$ is not necessarily equal to T_v or T_i. For the example considered in Eq. (1-46), $T_v = T_i = 2$, but the period of $p(t)$ is 1. If we denote the *minimum* period of $p(t)$ by T, then

$$p(t + nT) = v(t + nT)i(t + nT) = p(t) \tag{1-47}$$

In this case, it is more convenient to let $t_1 = nT$ and rewrite Eq. (1-45) in the equivalent form:

$$\begin{aligned}
P_{\text{av}} &= \lim_{n\to\infty} \frac{1}{nT} \int_0^{nT} v(t)i(t)\,dt \\
&= \lim_{n\to\infty} \frac{1}{nT} \left[\int_0^{T} v(t)i(t)\,dt + \int_T^{2T} v(t)i(t)\,dt \right. \\
&\qquad\qquad\qquad \left. + \cdots + \int_{(n-1)T}^{nT} v(t)i(t)\,dt \right] \\
&= \lim_{n\to\infty} \frac{1}{nT} \left[n \int_0^{T} v(t)i(t)\,dt \right] \\
&= \lim_{n\to\infty} \frac{1}{T} \int_0^{T} v(t)i(t)\,dt
\end{aligned}$$

Since the variable n no longer appears in this integral, the limit operation is superfluous and can be removed. Hence, *for periodic signals,* the average power can be written in the following simplified but equivalent form:

$$P_{\text{av}} = \frac{1}{T} \int_0^{T} v(t)i(t)\,dt = \frac{w(0,T)}{T} \tag{1-48}$$

where T is the minimum period of $v(t)i(t)$. Applying this formula to the same example considered in Eq. (1-46), we obtain

$$P_{\text{av}} = \frac{1}{1} \int_0^{1} (2 \sin \pi t) \left(\frac{2 \sin \pi t}{4} \right) dt = \frac{1}{2}$$

as we should.

Exercise: The voltage and current waveforms of a two-terminal black box are given, respectively, by $v = \sin (3.14)t$ and $i = \sin \pi t$. (*a*) Show that even though both $v(t)$ and $i(t)$ are periodic, the power $p(t)$ is not periodic. (*b*) For most practical purposes, $p(t)$ is said to be "almost periodic." Explain why.

The three expressions given by Eqs. (1-43), (1-45), and (1-48) are valid for any two-terminal black box. Let us now consider the special cases where the black box consists of a single nonlinear resistor, capacitor, or inductor. In so doing, we shall be able to derive a number of useful relationships. We shall also be able to draw some very important physical interpretations. Let us consider the three cases one at a time.

Case 1: Two-terminal nonlinear resistor Consider the nonlinear resistor shown in Fig. 1-21*a* and the three common types of v-i curves shown in Fig. 1-21*b*, *c*, and *d*. The v-i curve can be described in the functional form by $i = i(v)$ if it is voltage-controlled, or by $v = v(i)$ if it is current-controlled. A strictly monotonically increasing v-i curve can obviously be described by either $i = i(v)$ or $v = v(i)$. Accordingly, the instantaneous power flow $p_R(t)$, energy flow $w_R(t_0,t_1)$, and average power $P_{R_{\text{av}}}$ can be determined and are tabulated in Table 1-2 for these three cases.

Observe that corresponding to any operating point Q at any time t, the instantaneous power $p_R(t)$ is simply equal to the area of the shaded rectangles shown in Fig. 1-21. This power must, of course, come from the energy supplied by the external circuit connected across the resistor. From Table 1-2 we observe that the expressions for $p_R(t)$, $w_R(t_0,t_1)$, and $P_{R_{\text{av}}}$ depend on two pieces of information, namely,

1. The v-i curve
2. The voltage waveform $v(t)$ or the current waveform $i(t)$

Hence, in order to find out what happens to the power that enters the resistor, we must be given these two pieces of information. For

Fig. 1-21. The instantaneous power absorbed by a nonlinear resistor at any time t_0 is equal numerically to the area of the rectangle formed by the v,i axes and a vertex Q with coordinates $(v(t_0),i(t_0))$.

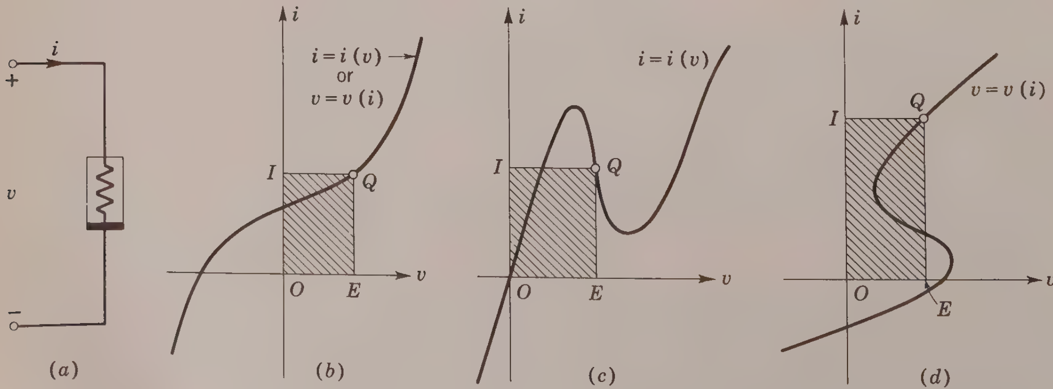


TABLE 1-2 *Instantaneous power, energy, and average power flow in a nonlinear resistor.*

	Strictly monotonically increasing v-i curve	Voltage-controlled v-i curve	Current-controlled v-i curve
$p_R(t)$	$= v(t)i(v(t))$ $= i(t)v(i(t))$	$= v(t)i(v(t))$	$= i(t)v(i(t))$
$w_R(t_0,t_1)$	$= \int_{t_0}^{t_1} v(t)i(v(t))\,dt$ $= \int_{t_0}^{t_1} i(t)v(i(t))\,dt$	$= \int_{t_0}^{t_1} v(t)i(v(t))\,dt$	$= \int_{t_0}^{t_1} i(t)v(i(t))\,dt$
$p_{R_{\text{av}}}$	$= \lim_{t_1\to\infty} \frac{1}{t_1}\int_0^{t_1} v(t)i(v(t))\,dt$ $= \lim_{t_1\to\infty} \frac{1}{t_1}\int_{t_0}^{t_1} i(t)v(i(t))\,dt$	$= \lim_{t_1\to\infty} \frac{1}{t_1}\int_0^{t_1} v(t)i(v(t))\,dt$	$= \lim_{t_1\to\infty} \frac{1}{t_1}\int_0^{t_1} i(t)v(i(t))\,dt$

example, suppose the v-i curve is represented by $i = v^3$, and the voltage is given by $v(t) = 2 \sin \pi t$. The instantaneous power can then be calculated; thus

$$p_R(t) = (2 \sin \pi t)(2 \sin \pi t)^3 = 16(\sin \pi t)^4$$

Observe that $p_R(t)$ has a period $T = 1$. The energy flow during the time interval $(0,t_1)$ and the average power due to the periodic signal are given, respectively, by

$$w_R(0,t_1) = 6t_1 - \frac{3}{\pi}\sin 2\pi t_1 - \frac{4}{\pi}(\sin \pi t_1)^3(\cos \pi t_1)$$

and

$$p_{R_{\text{av}}} = \frac{w_R(0,T)}{T} = \frac{w_R(0,1)}{1} = 6 \tag{1-49}$$

Equation (1-49) shows that even though the voltage $v(t)$ changes from positive to negative values periodically, there is a net positive average power flow entering the resistor. Since this power is not returned to the external circuit whenever the voltage returns to its initial value during each period, it cannot be recovered and is therefore said to be "lost" or "dissipated" in the resistor. Since energy cannot be destroyed, this loss of electrical energy in the resistor is merely transformed into heat energy.

The average power for the above example is positive. Let us now consider another example where this is not true. Suppose the

v-i curve is represented by $i = v^3 - 2$ and suppose a constant voltage $v = 1$ volt is applied. The instantaneous and average power, respectively, are given by

$$p_R(t) = 1(1-2) = -1 \qquad \text{and} \qquad p_{R_{\text{av}}} = \frac{(-1)(T)}{T} = -1 \tag{1-50}$$

Since the average power is negative, energy is being supplied (instead of being absorbed) by the nonlinear resistor to the external circuit. Since energy cannot be created, this nonlinear resistor must have an external power source (e.g., a battery) associated with it, and is therefore called an *active resistor*. Without an external power source, a nonlinear resistor can only absorb power; namely, $p_R(t) \geq 0$. Such a resistor is said to be *passive*. It is easy to see that *a nonlinear resistor is passive if, and only if, its v-i curve lies entirely in the first and the third quadrants*. This follows from the fact that the instantaneous power is always nonnegative; namely,

$$p_R(t) = v(t)i(t) \geq 0 \tag{1-51}$$

Clearly, in its original form, a physical resistor must necessarily be passive. This is true, for example, with the commercial resistors listed in Table 1-1. Any of these resistors can, of course, be transformed into an active resistor by connecting a battery in series with it.

Exercise: The v-i curve of a certain nonlinear resistor is given by $i = 10(v^3 - 3v)$ma, and the voltage excitation is given by $v(t) = 10 \sin t$ volts. (*a*) Find the instantaneous power $p_R(t)$. (*b*) Find the energy flow $w_R(0,t_1)$ for all $t_1 > 0$. (*c*) Find the average power by using Eq. (1-48) and check by using Eq. (1-45). (*d*) Is this nonlinear resistor passive or active? Explain why.

Case 2: Two-terminal nonlinear capacitor Consider the nonlinear capacitor shown in Fig. 1-22*a* and the three typical types of v-q curves shown in Fig. 1-22*b* to *d*. The v-q curve can be described in the functional form by $v = v(q)$ if it is charge-controlled or by $q = q(v)$ if it is voltage-controlled. A strictly monotonically increasing v-q curve can obviously be described by either $v = v(q)$ or $q = q(v)$. The energy flow $w_C(t_0,t_1)$ into a capacitor during the time interval (t_0,t_1) is given by

$$w_C(t_0,t_1) = \int_{t_0}^{t_1} v(t) \frac{dq(t)}{dt}\, dt \tag{1-52}$$

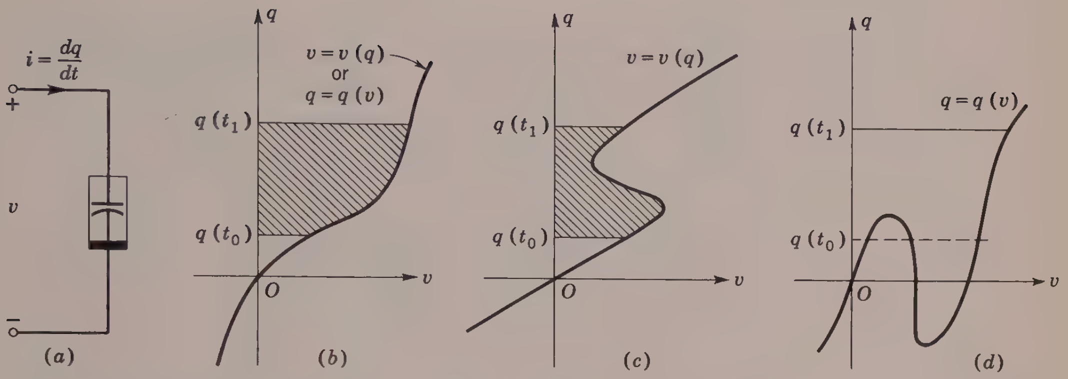


Fig. 1-22. The energy flow $w_C(t_0,t_1)$ from t_0 to t_1 into a nonlinear capacitor is equal numerically to the shaded area.

In the case where the capacitor v-q curve is either strictly monotonically increasing or charge-controlled, Eq. (1-52) can be written as

$$w_C(t_0,t_1) = \int_{t_0}^{t_1} v(q(t))\,\frac{dq(t)}{dt}\,dt \tag{1-53}$$

By a standard change of variable, Eq. (1-53) becomes

$$w_C(t_0,t_1) = \int_{q(t_0)}^{q(t_1)} v(q)\,dq \tag{1-54}$$

In the special, but very important, case of a *linear* capacitor [$q = Cv$ or $v = (1/C)q$], Eq. (1-54) can be reduced to

$$w_C(t_0,t_1) = \int_{q(t_0)}^{q(t_1)} \frac{1}{C}\,q\,dq = \frac{1}{2C}\int_{q(t_0)}^{q(t_1)} d(q^2)$$

or

$$w_C(t_0,t_1) = \frac{1}{2C}[q^2(t_1) - q^2(t_0)] \tag{1-55}$$

Equation (1-55) can also be expressed in terms of v by substituting $q = Cv$ for q:

$$w_C(t_0,t_1) = \frac{C}{2}[v^2(t_1) - v^2(t_0)] \tag{1-56}$$

Referring to Fig. 1-22, Eq. (1-54) can be interpreted as follows: The energy flow $w_C(t_0,t_1)$ from t_0 to t_1 into a charge-controlled nonlinear capacitor is equal numerically to the area under the v-q curve (bounded by the q axis and the lines $q = q(t_0)$ and $q = q(t_1)$. This interpretation is significant because it shows that only three pieces of information are needed to determine $w_C(t_0,t_1)$, namely

1. The v-q curve
2. The initial value of the charge at $t = t_0$
3. The final value of the charge at $t = t_1$

Since no information is required of the waveforms of $q(t)$ and $v(t)$, the energy $w_C(t_0,t_1)$ is said to be independent of the excitation waveforms. This property is very different from the resistor case where the complete voltage and current waveforms are required to compute $w_R(t_0,t_1)$. Observe further from Fig. 1-22 that whenever the waveform $v(t)$ returns to the same initial point, i.e., when $q(t_1) = q(t_0)$, the energy $w_C(t_0,t_1) = 0$. For example, Eqs. (1-55) and (1-56) are both equal to zero under this condition. Hence, unlike the resistor case, there must be some form of "energy-swapping" mechanism between a capacitor and the external circuit connected across it. To investigate this mechanism, let us calculate the average power using Eq. (1-45); thus

$$P_{C_{\text{av}}} = \lim_{t_1 \to \infty} \frac{1}{t_1} \int_{q(0)}^{q(t_1)} v(q)\, dq \tag{1-57}$$

Now observe that except when $q(t)$ goes to infinity, a case that cannot occur in practice, the value of $q(t_1)$ will always be a finite number. This means that the area under the curve representing the integral in Eq. (1-57) will always be a finite number. But the value of t_1 in Eq. (1-57) must tend to infinity, therefore

$$P_{C_{\text{av}}} = 0 \tag{1-58}$$

Since this equation is derived only under the assumption that the v-q curve be charge-controlled (this includes clearly the special case of a monotonically increasing curve), it is a very general result. We can, therefore, conclude that *the average power entering a charge-controlled nonlinear capacitor is zero.* This condition is true for any capacitor current and voltage waveforms. In the special case where $q(t)$ and $v(t)$ are periodic, Eq. (1-57) can be simplified to

$$P_{C_{\mathrm{av}}} = \frac{1}{T}\int_{q(0)}^{q(T)} v(q)\,dq \tag{1-59}$$

But $q(T) = q(0)$ for a periodic waveform of period T; therefore, Eq. (1-59) will integrate to zero, as it should.

From the preceding discussion, we can now conclude that a charge-controlled capacitor does not dissipate energy. Any energy entering it must be stored inside the capacitor and *may* eventually be returned. Because of this interpretation, a capacitor is often referred to as an *energy-storage element.* In the case of parallel-plate capacitors, it is possible to show, by electromagnetic field theory, that the energy is stored in the electric field between the plates. In view of this observation, the energy $w_C(t_0,t_1)$ in a capacitor is usually called the *electric stored energy.*

What happens if the v-q curve is neither monotonically increasing nor charge-controlled? In this case, it is no longer possible to describe the v-q curve by a function of q. It is not possible, therefore, to specify the area representing $v\,dq$ uniquely. To investigate this more general case, a new approach is required.[1]

Exercise 1: The v-q curve of a certain nonlinear capacitor is given by $q = \frac{1}{3} v^3$. Let the terminal voltage be given by $v(t) = e^{-t}$. (*a*) Find $w_C(0,t_1)$ for all $t_1 > 0$ by determining first $i(t) = (dq/dv)(dv/dt)$ and then using Eq. (1-43). (*b*) Repeat (*a*) by determining first $q(t)$ and then using Eq. (1-53). (*c*) Repeat (*a*) by using Eq. (1-54). (*d*) Let $v(t) = E \sin \omega t$ and verify that $P_{C_{\mathrm{av}}} = 0$.

Exercise 2: Prove that the electric stored energy in a voltage-controlled capacitor is given by

$$w_C(t_0,t) = q(t_1)v(t_1) - q(t_0)v(t_0) - \int_{v(t_0)}^{v(t_1)} q(v)\,dv$$

HINT: Apply the integration-by-part theorem.

Case 3: Two-terminal nonlinear inductor Consider the nonlinear inductor shown in Fig. 1-23*a* and the three typical types of i-φ curves shown in Fig. 1-23*b* to *d*. The i-φ curve can be described in the functional form by $i = i(\varphi)$ if it is flux-controlled or by $\varphi = \varphi(i)$ if it is current-controlled. A strictly monotonically increasing i-φ curve can obviously be described by either $i = i(\varphi)$ or $\varphi = \varphi(i)$. The energy flow $w_L(t_0,t_1)$ into an inductor during the time interval (t_0,t_1) is given by

$$w_L(t_0,t_1) = \int_{t_0}^{t_1} i(t)\,\frac{d\varphi(t)}{dt}\,dt \tag{1-60}$$

---

[1] This approach is called the *parametric approach* and is discussed in Appendix A. See also L. O. Chua and R. A. Rohrer, On the Dynamic Equations of a Class of Nonlinear *RLC* Networks, *IEEE Trans. Circuit Theory,* vol. CT-12, no. 4, pp. 475–489, December, 1965.

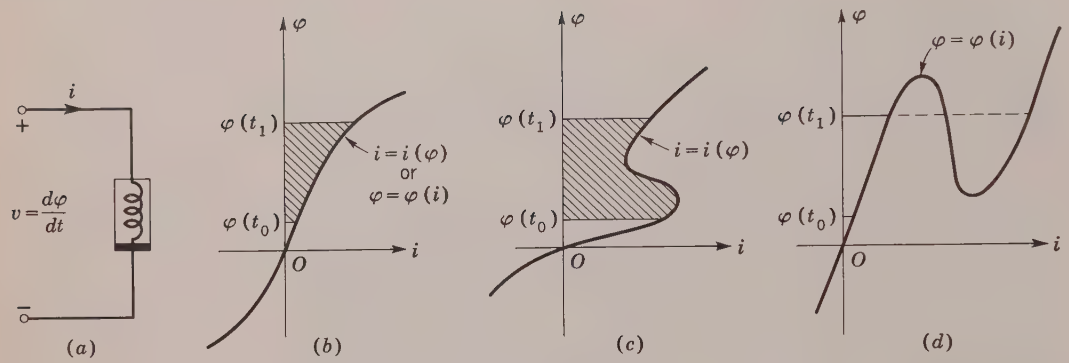


Fig. 1-23. The energy flow $w_L(t_0,t_1)$ from t_0 to t_1 into a nonlinear inductor is equal numerically to the shaded area.

Applying analogous procedure as in the capacitor case, we find that when the i-φ curve is either strictly monotonically increasing or flux-controlled, Eq. (1-60) can be written as

$$w_L(t_0,t_1) = \int_{\varphi(t_0)}^{\varphi(t_1)} i(\varphi)\, d\varphi \tag{1-61}$$

In the special case where the inductor is linear ($\varphi = Li$), Eq. (1-61) can be simplified further to

$$w_L(t_0,t_1) = \frac{1}{2L}[\varphi^2(t_1) - \varphi^2(t_0)] \tag{1-62}$$

or

$$w_L(t_0,t_1) = \frac{L}{2}[i^2(t_1) - i^2(t_0)] \tag{1-63}$$

Referring to Fig. 1-23, Eq. (1-61) can be interpreted as follows: The energy flow $w_L(t_0,t_1)$ from t_0 to t_1 into a flux-controlled nonlinear inductor is equal numerically to the area under the i-φ curve [bounded by the φ axis and the lines $\varphi = \varphi(t_0)$ and $\varphi = \varphi(t_1)$]. This interpretation has the same significance as for the capacitor; namely, only three pieces of information are needed to determine $w_L(t_0,t_1)$:

1. The i-φ curve
2. The initial value of the flux linkage at $t = t_0$
3. The final value of the flux linkage at $t = t_1$

By a similar procedure, we found *the average power in any flux-controlled inductor is zero;* thus

$$P_{L_{\text{av}}} = 0 \tag{1-64}$$

This means that a flux-controlled inductor cannot dissipate energy. In view of this observation, the inductor is also called an *energy-storage element.* In the case where the inductor is made of coils around an iron core, the energy can be shown, by electromagnetic principles, to be stored in the magnetic field around the coil. Hence, the energy stored in an inductor is usually called *magnetic stored energy.*

Exercise 1: Prove that Eq. (1-64) holds for a flux-controlled inductor. Verify this with $i(t) = I \cos \omega t$ and $\varphi = i^3$.

Exercise 2: Prove that the energy stored in a current-controlled inductor is given by

$$w_L(t_0,t_1) = \varphi(t_1)i(t_1) - \varphi(t_0)i(t_0) - \int_{i(t_0)}^{i(t_1)} \varphi(i)\, di$$

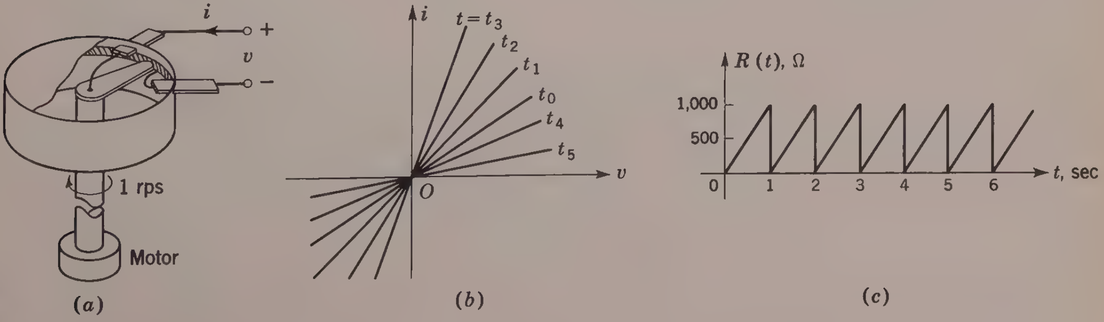
1-10 TIME-VARYING ELEMENTS

So far, the *v-i*, *v-q*, and *i-φ* curves characterizing a two-terminal resistor, capacitor, and inductor are assumed to remain unchanged for all times. These elements are said to be *time-invariant.* There exist some practical elements, however, whose *v-i*, *v-q*, or *i-φ* curves vary as functions of time. Such elements are said to be *time-varying resistors, capacitors,* or *inductors,* respectively.

Time-varying resistor The simplest example of a time-varying resistor is a potentiometer whose arm is being rotated by a motor as shown in Fig. 1-24*a*. At any time *t*, the potentiometer is simply a linear resistor with a straight-line *v-i* characteristic as shown in Fig. 1-24*b*. Hence, a time-varying linear resistor can be characterized by

$$v = R(t)i$$

Fig. 1-24. An example of a time-varying linear resistor.



where $R(t)$ is the time-varying resistance representing the reciprocal of the slope of the straight line at any time t. For example, if the potentiometer has a resistance range of 0 to 1,000 Ω uniformly distributed around its rim and if the arm rotates at a speed of 1 rps, then the time-varying resistance is as shown in Fig. 1-24*c*.

A time-varying resistor need not be linear. For example, consider a resistor characterized by

$$i = v^3 + \sin t$$

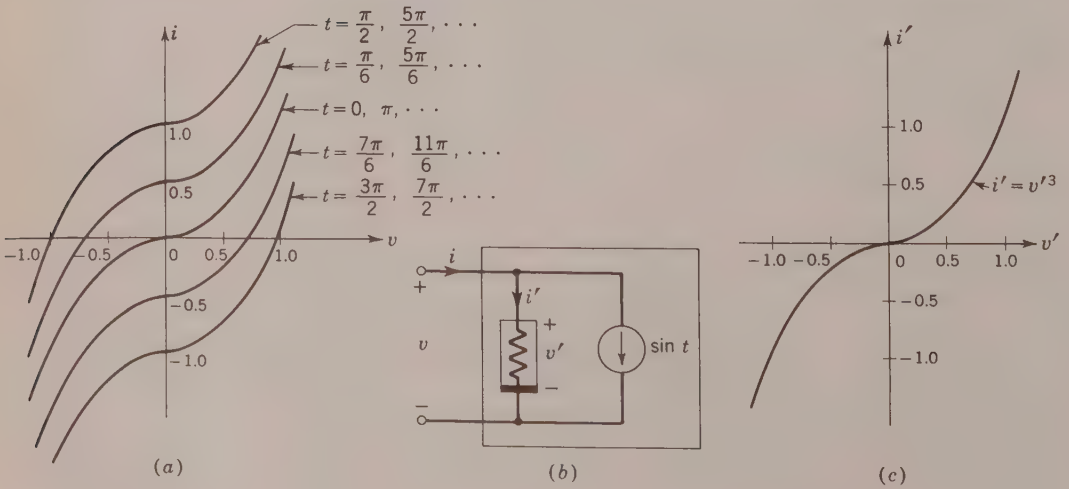
The v-i curve of this time-varying nonlinear resistor is shown in Fig. 1-25*a* as a function of time. Observe that this resistor can be constructed in practice by connecting a sinusoidal current source in parallel with a time-invariant resistor (Fig. 1-25*b*) with the v'-i' curve shown in Fig. 1-25*c*. In general, a time-varying nonlinear resistor can be characterized by a relationship $i = i(v,t)$ if it is voltage-controlled, or $v = v(i,t)$ if it is current-controlled. A review of the power and energy expressions derived in the preceding section would show that these expressions remain valid for the time-varying case.

What are time-varying resistors good for? To give one simple application, let us consider the current waveform

$$i(t) = [1 + f(t)] \sin \omega t \qquad (1\text{-}65)$$

Equation (1-65) is called an *amplitude-modulated waveform* because the amplitude of the sine wave varies with time. This is the type

Fig. 1-25. An example of a time-varying nonlinear resistor.



i
$t = \frac{\pi}{2}, \frac{5\pi}{2}, \ldots$
$t = \frac{\pi}{6}, \frac{5\pi}{6}, \ldots$
$t = 0, \pi, \ldots$
$t = \frac{7\pi}{6}, \frac{11\pi}{6}, \ldots$
$t = \frac{3\pi}{2}, \frac{7\pi}{2}, \ldots$
1.0
0.5
−1.0
−0.5
0
0.5
1.0
−0.5
−1.0
v
(a)
i
+
i'
+
v
v'
−
sin t
−
(b)
i'
1.0
0.5
$i' = v'^3$
v'
−1.0 −0.5
0
0.5
1.0
−0.5
−1.0
(c)

of signal that an AM radio transmitter sends out. In practice, $f(t)$ represents a slowly changing signal and $\sin \omega t$ represents a relatively high-frequency sine wave known as the "carrier." We are not equipped to explain why $f(t)$ cannot be transmitted directly, and why it must be "carried" by the sine wave. Suffice it to say that it takes a high-frequency waveform to traverse a long distance in space. Our objective here is to show how we may recover the signal $f(t)$ from Eq. (1-65). One possible method consists of applying this current to a time-varying linear resistor whose resistance changes at the *same* frequency as the carrier, namely,

$$R(t) = 1 + \sin \omega t$$

The voltage drop across this resistor is given by

$$\begin{aligned} v(t) &= R(t)i(t) \\ &= (1 + \sin \omega t)[1 + f(t)] \sin \omega t \\ &= \tfrac{1}{2} f(t) + \tfrac{1}{2} + [1 + f(t)] \sin \omega t - \tfrac{1}{2}[1 + f(t)] \cos 2\omega t \end{aligned} \tag{1-66}$$

Observe that Eq. (1-66) contains four terms; the first term is the signal that we would like to recover, the second term is a dc voltage, the third term is the carrier-frequency term, and the last term is at twice the carrier frequency. Through the use of a "filter," the last three components can be easily suppressed, thus leaving the desired signal $f(t)$. This recovering process is known as *synchronous detection* because the frequency of the time-varying resistance is synchronized at the same frequency as the carrier.

Exercise 1: Sketch the amplitude-modulated waveform given by Eq. (1-65) with $f(t) = \sin t$ and $\omega = 100$. What can you say about the "envelope" of this waveform?

Exercise 2: It is possible to rectify a sinusoidal current waveform $i(t) = I \sin t$ by applying this current to an appropriate time-varying linear resistance $R(t)$. Find $R(t)$ so that the resistor voltage is a rectified version of the current waveform; that is, $v(t) = i(t)$ whenever $i(t) \geq 0$, and $v(t) = 0$ whenever $i(t) \leq 0$.

Time-varying capacitor The simplest example of a time-varying linear capacitor is the air capacitor consisting of a fixed set of plates in mesh with a movable set of plates which is being rotated by a motor. A time-varying linear capacitor is therefore characterized by

$$q(t) = C(t)v(t) \tag{1-67}$$

where $C(t)$ is the time-varying capacitance. Unlike the resistor case, the expressions previously derived for the nonlinear capacitors do not apply in the time-varying case because when we differentiate $q(t)$ with respect to time, we obtain an additional term, namely,

$$i(t) = \frac{dq}{dt} = C(t)\frac{dv(t)}{dt} + v(t)\frac{dC(t)}{dt} \tag{1-68}$$

Since $C(t)$ is not a constant, the expressions given by Eqs. (1-55) and (1-56) are no longer applicable. Hence, to calculate the power or energy flow, we must resort to the original definitions.

Just as for the resistor, a time-varying capacitor may be nonlinear; in this case it is characterized by $q = q(v,t)$ if it is voltage-controlled or $v = v(q,t)$ if it is charge-controlled. Time-varying capacitors are useful in the study of parametric amplifiers. They are also useful in the modeling of many time-varying physical and biological systems. For example, the mass of a rocket during lift-off decreases rapidly with time as the rocket fuel is burned. This time-varying mass can be modeled by a time-varying capacitor.

Exercise 1: Find the average power $P_{C_{av}}$ entering a time-varying capacitor $C(t) = 2 - \cos \omega t$ and a terminal voltage $v(t) = E \sin \omega t$. Interpret whether this energy is being absorbed, delivered, or stored.

Exercise 2: Give an example for each of the following: (*a*) A time-varying linear capacitor, (*b*) a time-varying charge-controlled capacitor, and (*c*) a time-varying voltage-controlled capacitor.

Time-varying inductor By exact analogy to the capacitor, a time-varying linear inductor is characterized by

$$\varphi(t) = L(t)i(t) \tag{1-69}$$

where $L(t)$ is the time-varying inductance. Since $L(t)$ is no longer a constant, the expressions derived previously in the preceding sections are no longer valid. In particular, the inductor voltage is now given by

$$v(t) = \frac{d\varphi(t)}{dt} = L(t)\frac{di(t)}{dt} + i(t)\frac{dL(t)}{dt} \tag{1-70}$$

A time-varying inductor may be nonlinear; in this case it is characterized by $\varphi = \varphi(i,t)$ if it is current-controlled and $i = i(\varphi,t)$ if it is flux-controlled.

The analysis of a nonlinear network containing time-varying elements is a very difficult problem requiring advanced mathematics. Hence, we shall not consider any time-varying elements in the rest of this book. The above discussion is included mainly to emphasize the fact that most of the equations we derived in the previous sections are not valid for time-varying elements.

Exercise 1: Give an example for each of the following: (*a*) A time-varying linear inductor, (*b*) a time-varying flux-controlled inductor, and (*c*) a time-varying current-controlled inductor.

Exercise 2: Prove or disprove the assertion that if the current into a time-varying current-controlled inductor is periodic, then the instantaneous power $P_L(t)$ is also periodic.

1-11 CONCEPTS OF MODELING

One of the most basic principles in scientific analysis is that of modeling. Engineers and scientists seldom analyze a physical system in its original form. Instead, they construct a model which approximates the behavior of the system. By analyzing the behavior of the model, they hope to predict the behavior of the actual system. The primary reason for constructing models is that physical systems are usually too complex to be amenable to a practical analysis. In most cases, the complexity of a system is due in part to the presence of many nonessential factors. One basic principle of modeling consists, therefore, of extracting only the essential factors.

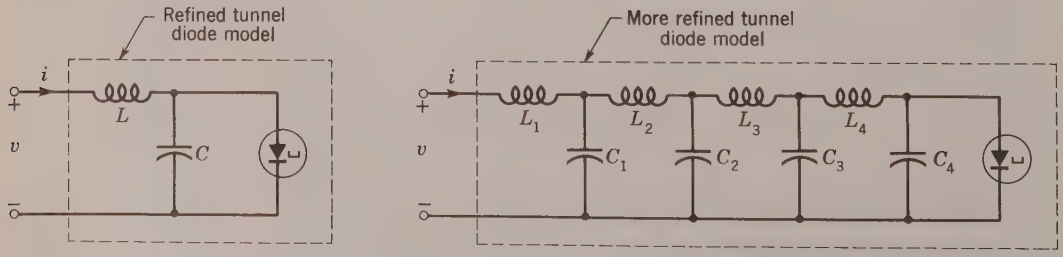
To illustrate the process of modeling, let us consider the problem of predicting the trajectory of a ballistic missile. This problem cannot be analyzed exactly because an exact analysis would require inclusion of all possible factors that may affect the trajectory. Some of these factors may be the weight and shape of the missile, the amount of thrust, the atmospheric drag, the deformation of the missile during flight, the distribution of weights of the internal components, the wind velocity, the impurity of the fuel, and the color of the missile. From experience, we know that the first three factors have a more significant influence on the trajectory than the remaining factors. This leads us to replace the missile by a model which includes only the first three factors. Obviously, the predicted trajectory based on this model is not going to be identical with that of the actual system. But as engineers, we are interested only in an "accurate" solution, not the exact solution. Hence, as long as the discrepancy between the predicted and the actual trajectories is tolerable, the model serves the purpose. Of course, in their desire to simplify analysis, engineers are often tempted to overidealize the model by stripping away some essential factors. In this case, the predicted solution may not be satisfactory.

The point we are driving at is that as engineers, we analyze the model which approximates an actual system. A model is always an idealization of a physical system. The more complex the model, the more accurate will the predicted solution be. Unfortunately, the analysis will also become more complicated. Hence, *a model is always a compromise between reality and simplicity.*

In the light of the above discussion, our definitions of a resistor, capacitor, and inductor must also be interpreted as models representing a physical device. For example, the v-i curve of the tunnel diode shown in Table 1-1 is a good model of a physical tunnel diode so long as the frequency at which we are operating is not very high. However, as the frequency increases, the static characteristic becomes less accurate, and a more realistic model must be found. For example, at very high frequencies, the connecting wires begin to behave like an inductance, and the capacitance between the wires gradually becomes significant. These elements are called "parasitic" or "stray" elements because they are invariably present, even though they are undesirable. A more realistic tunnel diode model must therefore include the parasitic elements such as the refined model shown in Fig. 1-26*a*. As the frequency gets higher, a still more complicated model such as shown in Fig. 1-26*b* may be chosen.

In this book, we shall be primarily interested in low-frequency models. In Chap. 11 we shall learn some basic techniques for constructing models of three-terminal devices in terms of two-terminal models. These low-frequency models can usually be refined for high-frequency analysis upon inclusion of appropriate parasitic elements.

Fig. 1-26. The static model of a tunnel diode must be refined for high-frequency analysis by the inclusion of appropriate parasitic inductances and capacitances at appropriate locations.



(*a*) (*b*)

1-12 SUMMARY

Four basic electrical variables

Charge q in coulombs

Flux linkage φ in webers

Current i in amperes

Voltage v in volts

Universal relationships

$$i = \frac{dq}{dt} \qquad q = \int_{-\infty}^{t} i(\tau)\, d\tau$$

$$v = \frac{d\varphi}{dt} \qquad \varphi = \int_{-\infty}^{t} v(\tau)\, d\tau$$

Simultaneity postulate The current entering one terminal of a two-terminal black box appears instantaneously at the other terminal.

Usual reference direction Current arrow enters the positive terminal or leaves the negative terminal. Under this assumption, the power $p(t)$ enters the black box whenever $p(t) > 0$.

Power and energy relationships

Instantaneous power: $p(t) = v(t)i(t)$

Energy flow from t_0 to t_1: $w(t_0,t_1) = \int_{t_0}^{t_1} v(\tau)i(\tau)\, d\tau$

Average power for arbitrary signal: $P_{\text{av}} = \lim_{t_1\to\infty} \frac{1}{t_1} \int_{t_0}^{t_1} v(t)i(t)\, dt$

Average power for periodic signal with period T:

$$P_{\text{av}} = \frac{1}{T} \int_{0}^{T} v(t)i(t)\, dt$$

Passive and active resistors A resistor is passive if, and only if, its v-i curve lies only in the first and the third quadrants. Otherwise, it is active.

Energy dissipation and storage properties A passive resistor can only dissipate energy, that is, $P_R(t) \geq 0$ for all time. This energy is lost as heat. A charge-controlled capacitor does not dissipate energy.

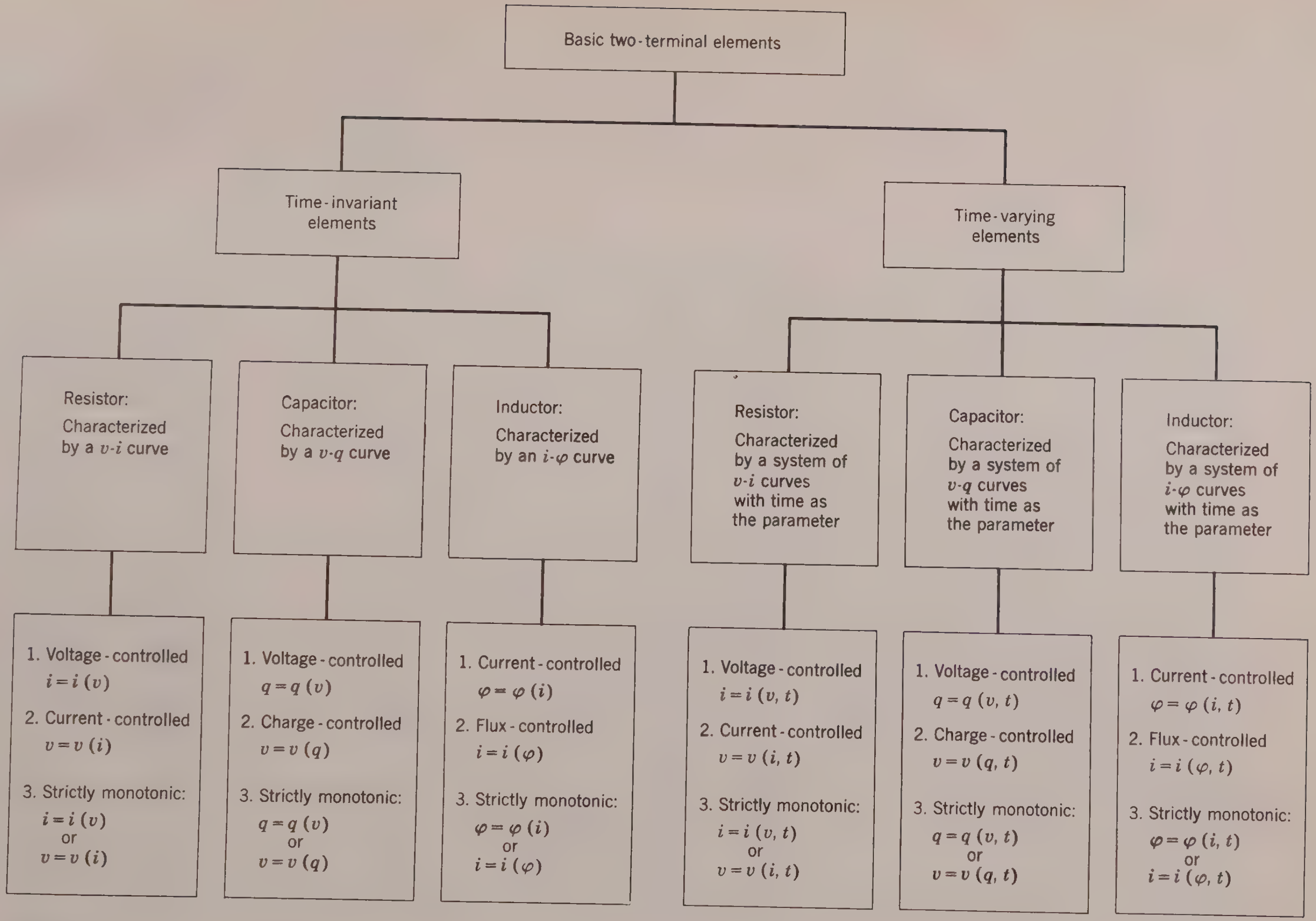

Basic two-terminal elements
Time-invariant elements
Time-varying elements
Resistor: Characterized by a v-i curve
Capacitor: Characterized by a v-q curve
Inductor: Characterized by an i-φ curve
Resistor: Characterized by a system of v-i curves with time as the parameter
Capacitor: Characterized by a system of v-q curves with time as the parameter
Inductor: Characterized by a system of i-φ curves with time as the parameter
1. Voltage-controlled $i = i(v)$ 2. Current-controlled $v = v(i)$ 3. Strictly monotonic: $i = i(v)$ or $v = v(i)$
1. Voltage-controlled $q = q(v)$ 2. Charge-controlled $v = v(q)$ 3. Strictly monotonic: $q = q(v)$ or $v = v(q)$
1. Current-controlled $\varphi = \varphi(i)$ 2. Flux-controlled $i = i(\varphi)$ 3. Strictly monotonic: $\varphi = \varphi(i)$ or $i = i(\varphi)$
1. Voltage-controlled $i = i(v, t)$ 2. Current-controlled $v = v(i, t)$ 3. Strictly monotonic: $i = i(v, t)$ or $v = v(i, t)$
1. Voltage-controlled $q = q(v, t)$ 2. Charge-controlled $v = v(q, t)$ 3. Strictly monotonic: $q = q(v, t)$ or $v = v(q, t)$
1. Current-controlled $\varphi = \varphi(i, t)$ 2. Flux-controlled $i = i(\varphi, t)$ 3. Strictly monotonic: $\varphi = \varphi(i, t)$ or $i = i(\varphi, t)$

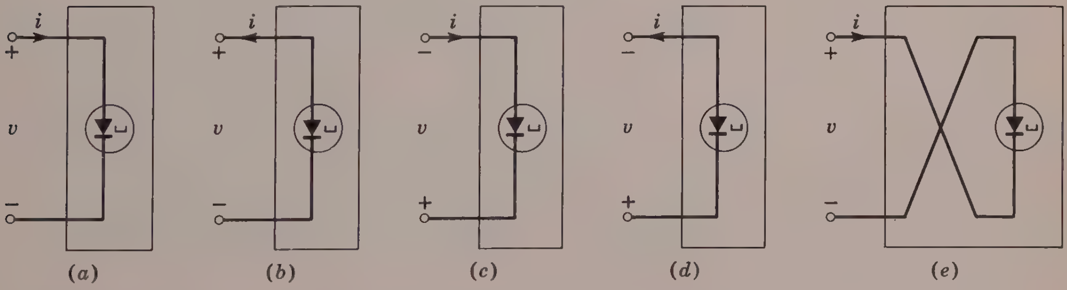
Fig. 1-27.

It stores *electric* energy and may be returned to its external circuit ($P_{C_{\rm av}} = 0$). A flux-controlled inductor does not dissipate energy. It stores *magnetic* energy and may be returned to its external circuit ($P_{L_{\rm av}} = 0$).

Basic two-terminal elements (See Fig. 1-27.)

PROBLEMS

1-1 In order to demonstrate the importance of reference direction and polarity, consider the following:
- (a) Sketch the v-i curve of each of the two-terminal black boxes shown in Fig. P1-1*a* to *e*. (The element inside the black box is a tunnel diode whose V-I curve is given in Table 1-1.)
- (b) Give a simple rule for sketching the v-i curve of a resistor whose terminals have been reversed as in Fig. P1-1*e*.
- (c) There exists a certain class of nonlinear elements in which it is unnecessary to distinguish between the two terminals. Such elements are called *bilateral elements.* All other elements are said to be *nonbilateral.* Give an example of a bilateral and a nonbilateral resistor, inductor, and capacitor.
- (d) Find the necessary and sufficient condition for a nonlinear resistor, capacitor, or inductor to be bilateral.

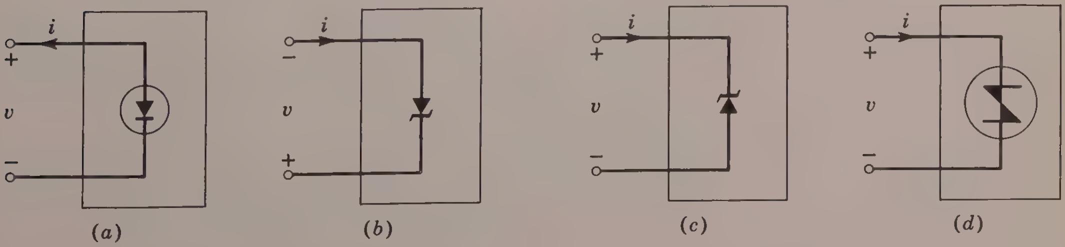


(*a*) (*b*) (*c*) (*d*) (*e*)

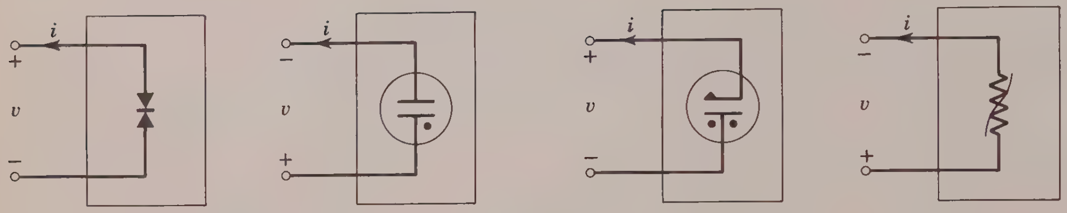
Fig. P1-1.

1-2 Sketch the v-i curve of each of the two-terminal black boxes shown in Fig. P1-2*a* to *h*. Refer to Table 1-1 for the V-I curves of the corresponding elements.

Fig. P1-2.



(*a*) (*b*) (*c*) (*d*)



(e) (f) (g) (h)

Fig. P1-2 (*Continued*).

1-3 The most accurate method for determining the v-i curve of a nonlinear resistor in the laboratory is the point-by-point method. Each point (E,I) on the curve is obtained by applying a voltage $v = E$ across the resistor and measuring the resulting current $i = I$. However, this method is rather tedious, and for most practical purposes it is desirable to design a v-i "curve tracer" that can display the v-i curve directly on an oscilloscope.

(a) Devise a simple circuit to carry out the above task using the ordinary 60-Hz ac line voltage and a Variac (variable voltage transformer) to provide the desired range of input voltage required by the given resistor. You may use a linear resistor whose voltage drop can be used to sense the magnitude of the current in the nonlinear resistor. To avoid grounding problems, you may use a 1:1 isolation transformer.

(b) Suppose that instead of using the line voltage as the energy source, we use the output voltage from a certain signal generator whose frequency can be changed from 10 Hz to 100 MHz. Do you expect the same v-i curve to be traced on the scope at all frequencies? If not, what frequency range must be chosen so that the v-i curve will agree approximately with the static curves supplied by the manufacturer?

1-4 It is sometimes convenient to define the dc resistance R_{dc} and the ac resistance R_{ac} at each point P of a v-i curve by

$$R_{dc} = \left.\frac{v}{i}\right|_P$$

$$R_{ac} = \left.\frac{dv}{di}\right|_P$$

(a) Show that $R_{dc} = \cot \alpha$, where α is the angle between the v axis and the straight line from the origin to point P.

(b) Show that $R_{ac} = \cot \beta$, where β is the angle between the v axis and the straight line tangent at point P.

(c) Verify that for a passive resistor, the value of R_{dc} is always positive.

(d) Verify that the value of R_{ac} may be either negative, zero, positive, or even infinite for a passive resistor.

(e) Sketch the relationships R_{dc} versus v and R_{ac} versus v for the varistor type 1NXX5, the zener diode type 1NXX3, and the tunnel diode type 1NXX6. See Appendix D for the v-i curve of these elements.
(f) Repeat (*e*) for the relationships R_{dc} versus i and R_{ac} versus i.

1-5 Consider the definitions of the dc resistance R_{dc} and ac resistance R_{ac} given in Problem 1-4.
(a) If the resistor v-i curve is strictly monotonically increasing, what can you say about the curves R_{dc} versus v, R_{dc} versus i, R_{ac} versus v, and R_{ac} versus i? Are they monotonic, single-valued, or multivalued?
(b) Repeat (*a*) for a voltage-controlled resistor.
(c) Repeat (*a*) for a current-controlled resistor.

1-6 It is sometimes convenient to define the dc conductance G_{dc} and the ac conductance G_{ac} at each point P of a v-i curve by

$$G_{dc} = \left.\frac{i}{v}\right|_P$$

$$G_{ac} = \left.\frac{di}{dv}\right|_P$$

(a) Show that $G_{dc} = \tan \alpha$, where α is the angle between the v axis and the straight line from the origin to point P.
(b) Show that $G_{ac} = \tan \beta$, where β is the angle between the v axis and the straight line tangent at point P.
(c) Verify that for a passive resistor, the value of G_{dc} is always a finite, positive number.
(d) Verify that the value of G_{ac} may be either negative, zero, positive, or even infinite for a passive resistor.
(e) Sketch the relationships G_{dc} versus i and G_{ac} versus i for the varistor type 1NXX5, the zener diode type 1NXX3, and the tunnel diode type 1NXX6. See Appendix D for the v-i curve of these elements.
(f) Repeat (*e*) for the relationships G_{dc} versus v and G_{ac} versus v.

1-7 Consider the definitions of the dc conductance G_{dc} and ac conductance G_{ac} given in Prob. 1-6.
(a) If the resistor v-i curve is strictly monotonically increasing, what can you say about the curves G_{dc} versus i, G_{dc} versus v, G_{ac} versus i, and G_{ac} versus v? Are they monotonic, single-valued, or multivalued?
(b) Repeat (*a*) for a voltage-controlled resistor.
(c) Repeat (*a*) for a current-controlled resistor.

1-8 If a sinusoidal voltage waveform $v = A \sin \omega t$ is applied across a nonlinear resistor characterized by a polynomial

$$i = a_0 + a_1 v + a_2 v^2 + \cdots + a_n v^n$$

the resulting current will contain, in addition to the fundamental frequency term $i = B \sin \omega t$, other higher harmonic terms. In many practical applications, these harmonic terms are usually filtered out, in which case it becomes meaningful to define the ratio between the amplitudes of the fundamental voltage and current components to be the average resistance R_{av}.

(a) Find the average resistance R_{av} with $n = 3$.
(b) Assuming $a_0 = a_1 = a_2 = a_3 = 1$, plot R_{av} versus the amplitude A.
(c) What is the significance of the R_{av}-vs.-A curve obtained in (*b*)?

1-9 If a sinusoidal voltage waveform $v = A \cos \omega t$ is applied across a voltage-controlled capacitor characterized by

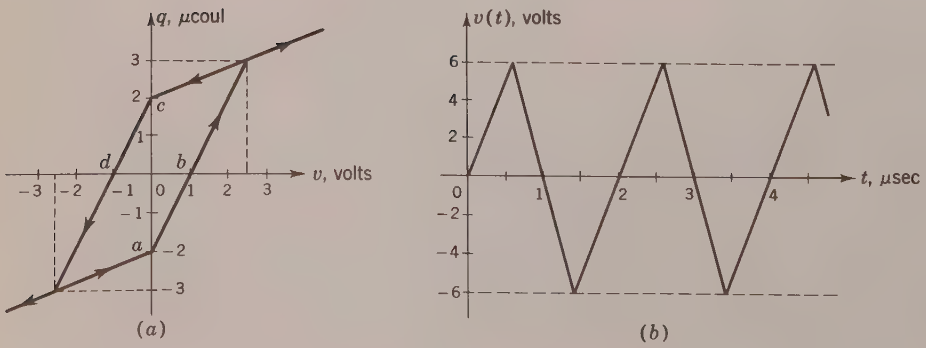
$$q = a_0 + a_1 v + a_2 v^2 + \cdots + a_n v^n$$

the resulting current will contain, in addition to the fundamental frequency term $i = B \sin \omega t$, other higher harmonic terms. In many practical applications these harmonic terms are filtered out, in which case the ratio between the amplitudes of the fundamental voltage and current components is usually called the *describing function* Z_C.

(a) Find the describing function Z_C with $n = 3$.
(b) Observe that unlike the average resistance in Prob. 1-8, the describing function Z_C of a capacitor is a function of the frequency ω. Assuming $a_0 = a_1 = a_2 = a_3 = A = 1$, plot the curve Z_C versus ω.
(c) What is the significance of the Z_C-vs.-ω curve?

1-10 The v-q curve of a practical, nonlinear capacitor with a barium titanate dielectric is shown in Fig. P1-10*a*. If the triangular voltage signal shown in Fig. P1-10*b* is applied across this capacitor, find the current waveform $i(t)$. Assume that the capacitor is operating initially at point *a* of the hysteresis curve. Assume also that the locus must follow the arrow directions.

Fig. P1-10.

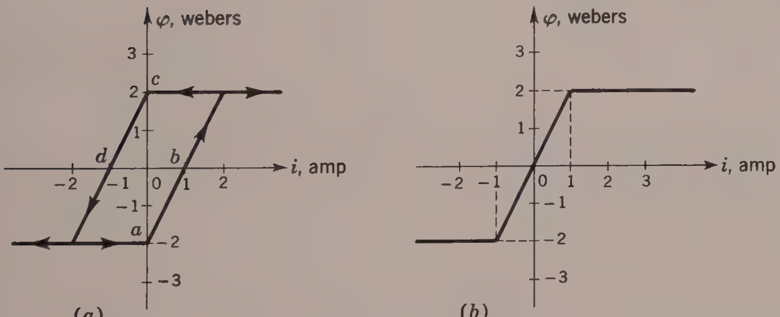


1-11 Based on the definition in Prob. 1-9, define an analogous *describing function* Z_L for a current-controlled inductor. Plot the Z_L-vs.-ω curve for the following cases (assume $A = 1$):

(a) $\varphi = i$
(b) $\varphi = i^2$
(c) $\varphi = i^3$
(d) $\varphi = 1 + i + i^2 + i^3$

1-12 The i-φ curve of an iron-core nonlinear inductor is usually characterized by a hysteresis loop as shown in Fig. P1-12*a*.

(a) If a sinusoidal current $i(t) = 8 \sin t$ amp is applied to this inductor, find the voltage $v(t)$ for $t \geq 0$ and the average power P_{av}. (Assume that the inductor is operating at point a at $t = 0$.)
(b) Does this inductor dissipate or store energy?
(c) Verify that the average power P_{av} is equal numerically to the area enclosed by the hysteresis loop.
(d) To simplify analysis, we often approximate a hysteresis loop by a single-valued curve, such as the i-φ curve shown in Fig. P1-12*b*. Repeat parts (*a*), (*b*) and (*c*) for this case and compare the solutions.



(*a*) (*b*)

Fig. P1-12.

1-13 For economic reasons, it is desirable to transmit more than one signal simultaneously over the same communication channel such as a long-distance telephone system. To be specific, let us suppose that two voice signals $f_1(t)$ and $f_2(t)$ are to be transmitted from Chicago to Paris. Since the highest frequency of the human voice is about 20 kHz, it is not possible to transmit these signals directly (it can be shown by electromagnetic field principles that it takes a very high-frequency signal to propagate in space). In practice, our transmitter in Chicago can be designed to combine these signals into the following "amplitude-modulated" current waveform

$$i(t) = f_1(t) \sin \omega t + f_2(t) \cos \omega t$$

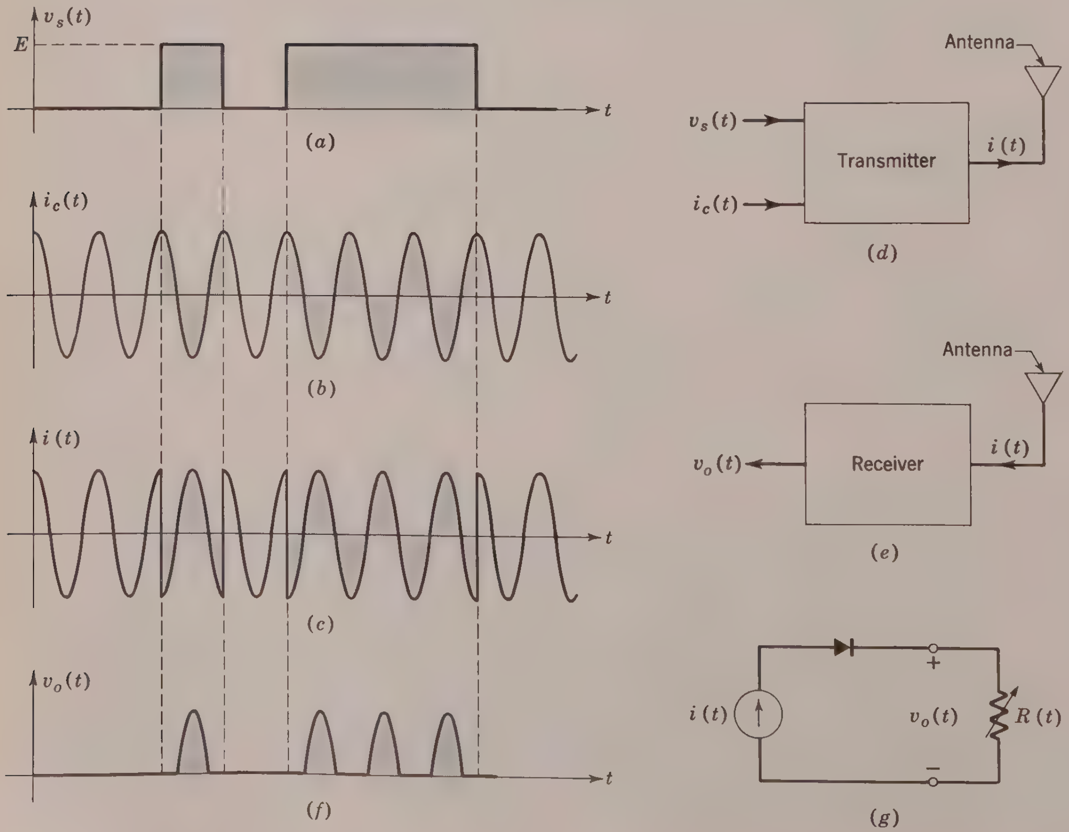
where $\sin \omega t$ and $\cos \omega t$ are very high-frequency (say 10 MHz) sine waves. This is the signal that will be received in Paris. Our problem is to recover $f_1(t)$ and $f_2(t)$ at the receiving end.

(a) Show that $f_1(t)$ can be recovered by applying $i(t)$ to a time-varying linear resistor with $R(t) = 1 + \sin \omega t$ and then suppressing the components with a frequency higher than ω by means of a filter.

(b) By a similar procedure, $f_2(t)$ can be recovered. Find the appropriate time-varying resistance $R(t)$ for accomplishing this.

1-14 One method for transmitting a telegraph signal over long distances is to modulate the "phase" of a high-frequency sinusoidal signal called the *carrier*. To be specific, suppose we wish to transmit the letter A in morse code by closing and opening the telegraph key at appropriate intervals. Corresponding to this code, the output voltage $v_s(t)$ shown in Fig. P1-14*a* will be generated. In order to transmit this waveform over long distances, an apparatus can be

Fig. P1-14.



designed to change the phase of the carrier signal $i_c(t)$ shown in Fig. P1-14*b* abruptly by 180° each time $v_s(t)$ changes its amplitude. For example, the resulting current waveform $i(t)$ is shown in Fig. P1-14*c*. This is the signal being transmitted and received, as shown in Fig. P1-14*d* and *e*. Our problem is to decode the received current waveform $i(t)$ so as to recover the message A. This can be accomplished by applying $i(t)$ (as a current source) to the time-varying circuit shown in Fig. P1-14*g* so as to produce the output voltage $v_o(t)$ shown in Fig. P1-14*f*. Observe that even though $v_o(t)$ is not identical with $v_s(t)$, the nature of the waveform is unmistakably similar to that of Fig. P1-14*g*. Hence the above decoding scheme would accomplish our objective.

(a) Specify the time-varying resistance $R(t)$ for accomplishing this task.

(b) What can you say about the frequency of $R(t)$ in comparison with that of the carrier signal $i_c(t)$?

(c) The above scheme for decoding the signal is known as the *synchronous phase detection.* Explain why.

2 CONTROLLED ELEMENTS

2-1 TWO-TERMINAL CONTROLLED ELEMENTS

There exist many practical two-terminal resistors, capacitors, and inductors whose v-i, v-q, and i-φ curves depend on some physical variable x. Such elements are called *controlled resistors, controlled capacitors,* and *controlled inductors,* respectively. The independent variable x is called the *controlling* variable and may represent a nonelectrical variable such as temperature, light intensity, pressure, force, or velocity, or an electrical variable such as voltage and current.[1] Actually, to be exact, all physical two-terminal elements are controlled elements in the above sense because every physical device is affected, to a certain extent, by temperature changes. In most cases, however, the temperature variation is undesirable, and one of the goals of the manufacturer is to see to it that the temperature effect is minimized. In contrast with this, the controlled elements that we are studying in this chapter are specifically designed so that the controlling variable x produces a significant effect on the characteristic curves of these elements. In fact, we shall try to take advantage of the variation of the characteristic curves to design practical circuits.

2-1-1 TWO-TERMINAL CONTROLLED RESISTORS

A two-terminal *controlled resistor* is characterized by a family of v-i curves, each curve corresponding to a specific value of the controlling variable x. A controlled resistor may be represented mathematically by $i = i(v,x)$ if it is voltage-controlled for all values of x, or $v = v(i,x)$ if it is current-controlled for all values of x. A strictly monotonic controlled resistor can be represented in either form.

---

[1] The time-varying elements defined in Sec. 1-10 may also be interpreted as a controlled element with the time t as the controlling variable. However, unlike the other physical variables, we have no control over the value of t. Hence, we shall exclude time-varying elements from the context of controlled elements.

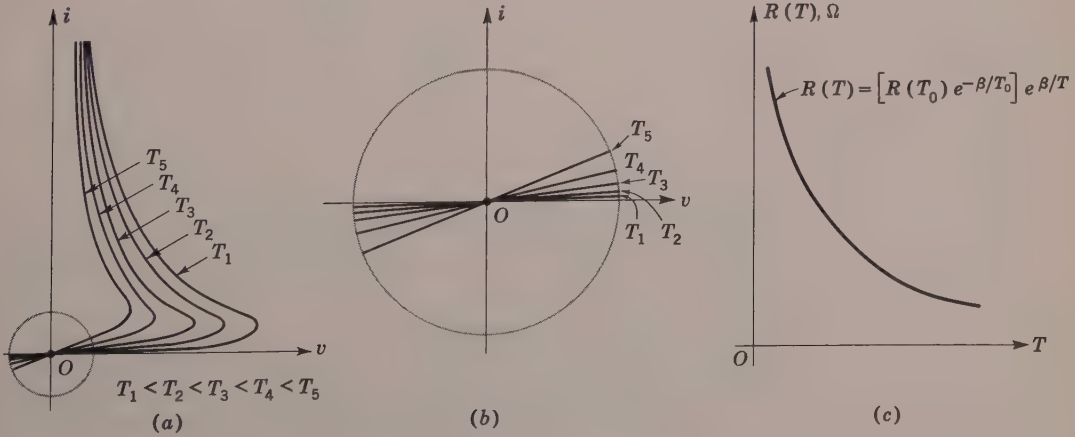
A number of commercially available controlled resistors are listed in Table 2-1. The first two elements are called *thermistors* because the controlling variable is temperature. The next three elements (the *phototube,* the *photodiode,* and the *solar cell*) have light intensity as the controlling variable. The last element listed is called a *cryotron* and is controlled by an external current I_C, independent of the terminal variables I and V.

Observe that the v-i curves of the controlled resistors listed in Table 2-1 are all nonlinear. With the exception of the solar cell, all elements listed in Table 2-1 are characterized by a family of v-i curves in the first and the third quadrants only (in most cases, only curves in the first quadrant are shown). Therefore, these elements are passive, nonlinear, controlled elements. The v-i curves of the solar cell fall in the fourth quadrant, and it is therefore an active, nonlinear, controlled element. We do not violate the conservation of energy here because the solar cell does have a source of energy, solar energy.

In many practical applications, we are interested in operating only in the linear region of the v-i curves. For example, consider the NTC thermistor curves near the origin (Fig. 2-1a). The v-i curves in this region are enlarged as shown in Fig. 2-1b. Observe that these are straight lines, and hence the v-i curves can be represented mathematically by

$$v = R(T)i \tag{2-1}$$

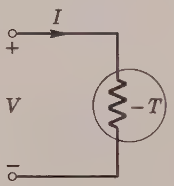
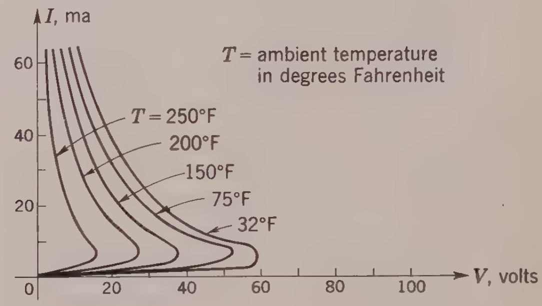
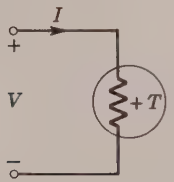
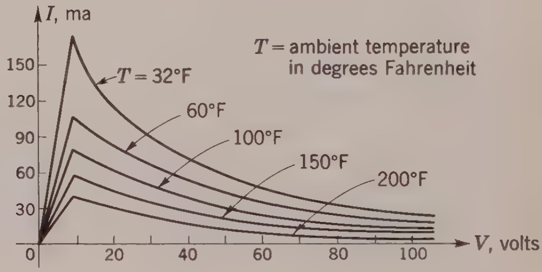
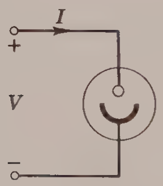
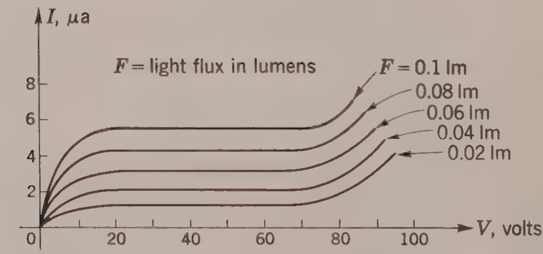
where $R(T)$ is the resistance of each curve at temperature T. For the NTC thermistor, the resistance-vs.-temperature curve assumes



(a) (b) (c)

Fig. 2-1. An NTC thermistor, when operated around the vicinity of the origin, can be considered as a linear controlled resistor and can, therefore, be characterized by a resistance-vs.-temperature curve.

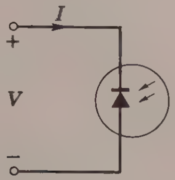
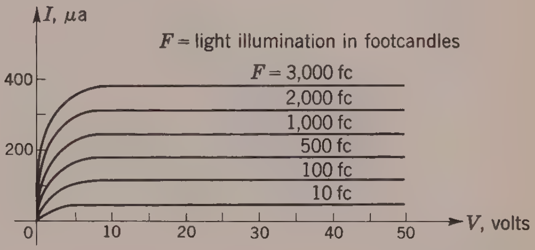
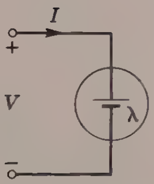
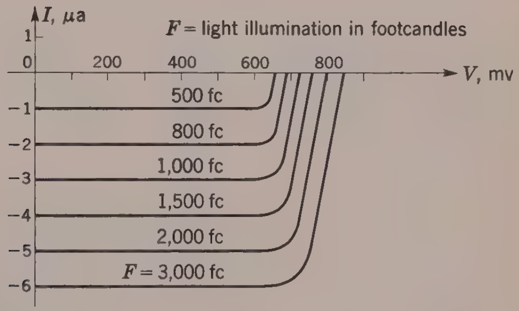
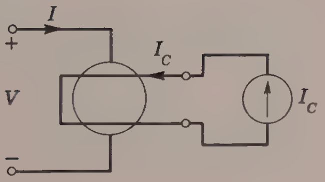
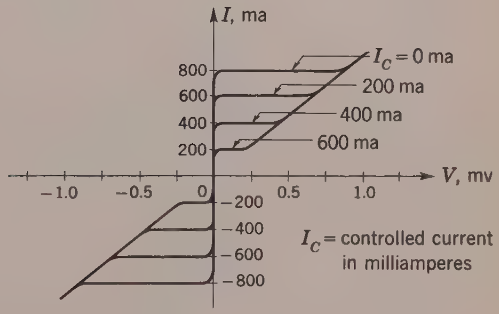
TABLE 2-1 *Practical two-terminal controlled resistors.*

<table>
<tr><th>Name</th><th>Symbol</th><th><i>v-i</i> characteristic curves</th></tr>
<tr><td>NTC (negative-temperature-coefficient) thermistor</td><td>
</td><td>
</td></tr>
<tr><td>PTC (positive-temperature-coefficient) thermistor</td><td>
</td><td>
</td></tr>
<tr><td>Phototube</td><td>
</td><td>
</td></tr>
</table>

the general form shown in Fig. 2-1*c*. For this simple curve, the following mathematical function has been found to be a good approximation:

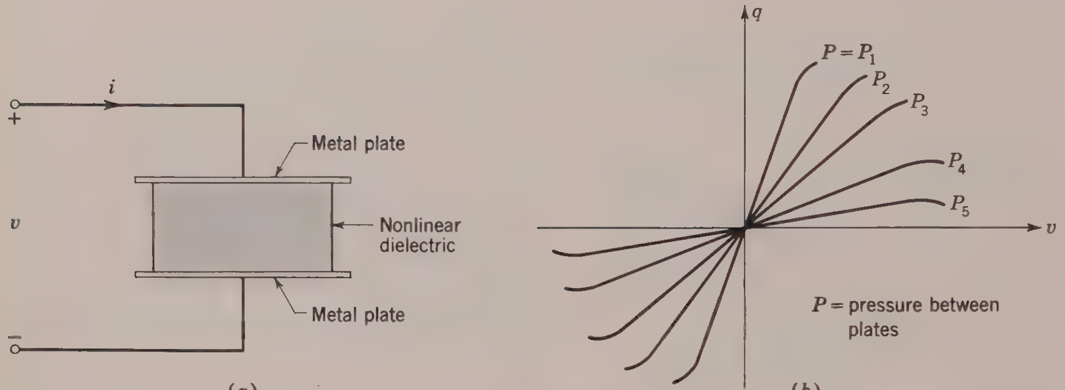
$$R(T) = [R(T_0)e^{-\beta/T_0}]e^{\beta/T} \tag{2-2}$$

where $R(T_0)$ is the resistance measured at some reference tem-

me	Symbol	*v-i* characteristic curves
otodiode		
lar cell hotocell)		
yotron		

perature T_0 (usually 298°K), and β is a constant which depends on the material of which the thermistor is made.

Exercise 1: The parameters of a certain commercial NTC thermistor were measured as follows: $R(T_0) = 10\text{k}\ \Omega$, $\beta = 3{,}000$, $T_0 = 298°\text{K}$. (*a*) Sketch the $R(T)$ curve as defined by Eq. (2-2) for the temperature range $300°\text{K} < T < 500°\text{K}$. (*b*) Sketch the *v-i* curve with T in degrees Celsius as the parameter.



(a) (b)

Fig. 2-2. A pressure-controlled nonlinear capacitor is characterized by a family of v-q curves, each curve corresponding to some applied pressure.

Exercise 2: It is convenient to define a temperature coefficient α for $R(T)$ by the formula

$$\alpha = \frac{1}{R(T)} \frac{dR(T)}{dT}$$

(*a*) Show that for the NTC thermistor, $\alpha = -\beta/T^2$. (*b*) Show that α can be interpreted to be the percentage change in resistance per degree temperature change. (*c*) Calculate the α of the thermistor given in Exercise 1. Why do we call this a negative-temperature coefficient thermistor?

Exercise 3: The resistance $R(T)$ of a PTC thermistor operating at the low-voltage region is given approximately by

$$R(T) = A + C\,e^{BT}$$

where A, B, and C are positive constants and T is the temperature in degrees Fahrenheit. (*a*) Sketch the general shape of the curve $R(T)$. (*b*) Show that the temperature coefficient α (see Exercise 2) is given by

$$\alpha = \frac{BC\,e^{BT}}{A + C\,e^{BT}}$$

Exercise 4: The family of v-i curves of a photodiode can be represented approximately by

$$i = I(F) + I_o(1 - e^{-kv})$$

where I_o and k are constants and $I(F)$ is a function of the light illumination F. (*a*) Find the approximate values of I_o and k, and sketch the curve $I(F)$ for the photodiode type 1NXX10 given in Appendix D. (*b*) Show that when $I(F) = 0$, the photodiode reduces to a junction diode [see Eq. (1-14) of Chap. 1] with its terminals interchanged.

Exercise 5: The family of v-i curves of a solar cell can be represented approximately by

$$i = I_o(e^{kv} - 1) - I(F)$$

where I_o and k are constants and $I(F)$ is a function of the light illumination F. (*a*) Find the approximate values of I_o and k, and sketch the curve $I(F)$ for the solar cell type 1NXX11 given in Appendix D. (*b*) Show that when $I(F) = 0$, the solar cell reduces to a junction diode [see Eq. (1-14) of Chap. 1].

2-1-2 TWO-TERMINAL CONTROLLED CAPACITORS

A two-terminal *controlled capacitor* is characterized by a family of v-q curves, each curve corresponding to a specific value of the controlling variable x. A controlled capacitor may be represented mathematically by $q = q(v,x)$ if it is voltage-controlled, or $v = v(q,x)$ if it is charge-controlled. A strictly monotonic controlled capacitor can be represented in either form.

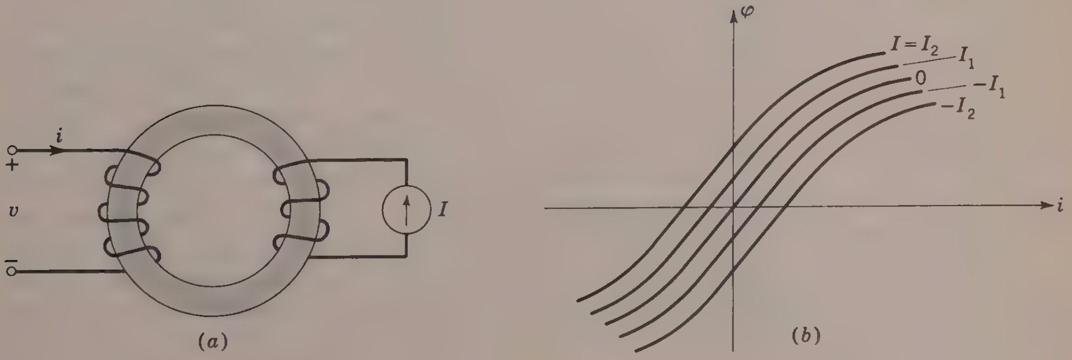
Although there are not many controlled capacitors presently available in commercial quantities, they do exist in various forms for special applications. For example, one can easily build a pressure-controlled nonlinear capacitor by sandwiching a nonlinear dielectric material between a pair of metallic plates as shown in Fig. 2-2*a*. For each value of applied pressure, a different v-q curve is measured as shown in Fig. 2-2*b*.

2-1-3 TWO-TERMINAL CONTROLLED INDUCTORS

A two-terminal *controlled inductor* is characterized by a family of i-φ curves, each curve corresponding to a specific value of the controlling variable x. A controlled inductor may be represented mathematically by $\varphi = \varphi(i,x)$ if it is current-controlled, or $i = i(\varphi,x)$ if it is flux-controlled. A strictly monotonic controlled inductor can be represented in either form.

An example of a practical, controlled inductor is the toroidal coil shown in Fig. 2-3*a*. By applying a direct current through a separate coil, the i-φ curve can be translated along a vertical direction as shown in Fig. 2-3*b*.

Fig. 2-3. A toroidal coil becomes a current-controlled nonlinear inductor by applying a dc current through a separate winding.



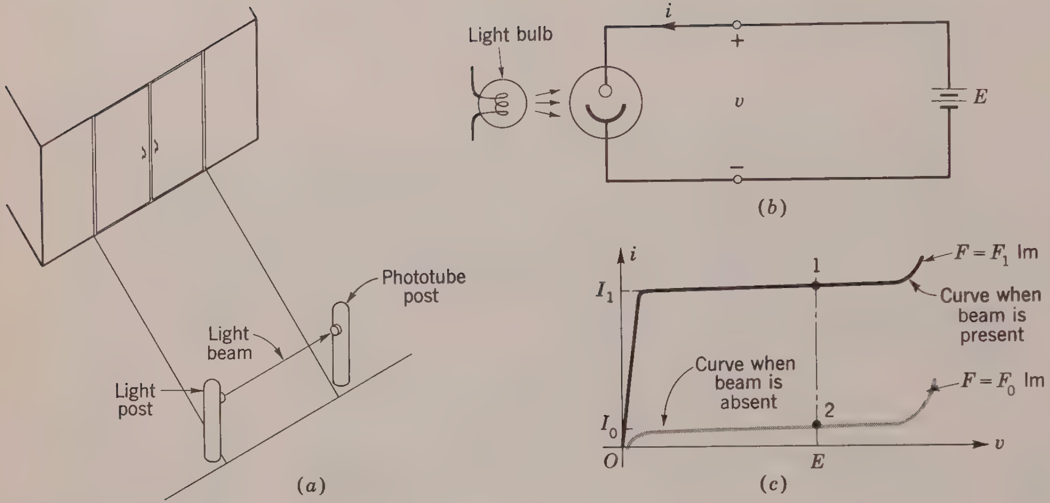


Fig. 2-4. The basic door-opening mechanism consists of a phototube circuit and a beam of light.

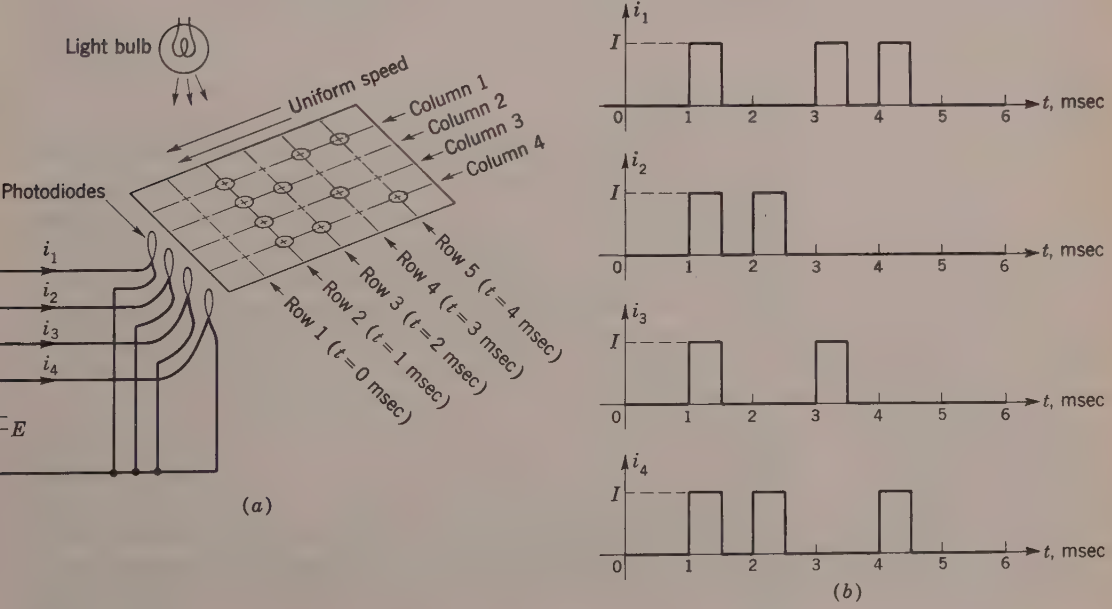
2-2 PRACTICAL APPLICATIONS OF CONTROLLED ELEMENTS

Controlled elements are widely used in the instrumentation and control industries. In almost all cases, they are used for "sensing" or "detection" purposes. For example, the familiar electric eye for actuating an automatic door-opening mechanism consists simply of a phototube and a beam of light as shown in Fig. 2-4*a*. For simplicity, let us connect a battery across the phototube as shown in Fig. 2-4*b*. The light bulb is positioned so that the light beam produces a light illumination of x lm on the phototube at all times, except when it is interrupted by a person walking past the doorway. In the latter case, the light illumination at the phototube will be decreased to an insignificant value such as F_0. Accordingly, the v-i curve of the phototube will consist of either one of the two curves shown in Fig. 2-4*c*. Corresponding to an applied voltage E, a current $i = I_1$ will be measured if the beam is present. However, if the beam is cut off, the current will drop back to a small value $i = I_0$. Hence, if we connect a sensitive relay in series with the phototube, then the relay contact, which is previously "open," will become "closed" whenever the current drops to I_0, thereby deenergizing the relay. The closing of the contacts will serve as a switch to turn on the motor for opening the door. In practice, phototube circuits are, of course, much more complicated, and we shall not be able to analyze them until after Chap. 6.

Another common application of a controlled resistor is the punch-card-reader mechanism shown in Fig. 2-5*a*. In view of its smaller size, the photodiode is chosen in favor of the phototube. Our objective is to transform the code represented by the presence or absence of holes in the computer card into a corresponding sequence of electrical signals. By installing a string of identical photodiodes below the corresponding columns, it is easy to see that the circuit shown in Fig. 2-5*a* will indeed generate the desired signals. To be specific, let us assume that the card has four columns and five rows. Let us also assume that the card is being fed at a uniform speed so that row 1 will appear above the photodiodes at $t = 0$ msec, row 2 at $t = 1$ msec, . . . , and row 5 at $t = 4$ msec. It is easy to see that the waveforms of i_1, i_2, i_3, and i_4 will be as shown in Fig. 2-5*b*.

There are many more practical applications of this nature. In fact, two-terminal controlled resistors are widely used in applications which require the detection or sensing of temperature or light-illumination variations. For example, thermistors have been used for designing fire-protection devices, pyrometers, flowmeters, vacuum gages, etc. Two-terminal controlled capacitors are useful in applications which require the detection of small pressures, forces, or distances. For example, nonlinear controlled capacitors

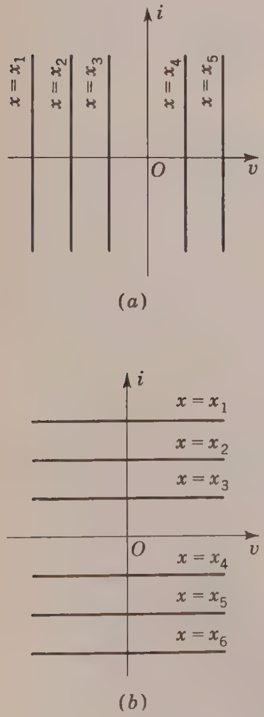
Fig. 2-5. A punch-card reader consisting of a string of photodiodes for sensing the light illumination and hence the presence or absence of a hole.



using barium titanate dielectrics have been used extensively in the design of strain gages. Two-terminal controlled inductors are widely used in connection with those applications requiring an external controlled current or magnetic-field intensity. For example, they are used extensively in the control of lighting effects in theaters and auditoriums.

One could go on and list many other interesting applications. However, there is really only one concept involved in all these applications. We ask ourselves first the question: "What is the physical variable to be controlled?" Once the variable x is chosen, we look for an appropriate controlled resistor having x as the controlling variable. We then design a basic circuit to transform the variations in x into corresponding variations in voltage or current. Finally, we try to optimize the circuit by introducing additional elements at appropriate locations. We shall be able to do all these after studying the materials in this book.

Fig. 2-6. The v-i curves of an x-controlled voltage source consist of a family of vertical lines, whereas those of an x-controlled current source consist of a family of horizontal lines.



Exercise 1: Discuss how a thermistor might be used in the design of the following gadgets: (*a*) thermometers; (*b*) liquid-level gages for indicating the level of liquids inside a closed tank; (*c*) flowmeters for measuring the number of gallons per second of liquids flowing through a pipe; (*d*) fire alarm indicators; (*e*) gas leakage detectors; (*f*) vacuum gages.

Exercise 2: Discuss how a solar cell might be used in the design of the following gadgets: (*a*) photographic exposure meters; (*b*) automatic street lighting control; (*c*) pinhole or flaw detector in sheet metal or paper manufacturing; (*d*) remote flashlight switch for turning on and off a television set.

2-3 CONTROLLED SOURCES

We shall now define a special class of two-terminal controlled resistors characterized either by a family of vertical v-i curves (Fig. 2-6*a*) or a family of horizontal v-i curves (Fig. 2-6*b*). Since for each value of x the v-i curve reduces to that of a voltage source in the vertical case and that of a current source in the horizontal case, these elements are called *x-controlled voltage sources* and *x-controlled current sources*, respectively. It is important to observe that the v-i curves of an x-controlled voltage source are independent of the terminal current i, and can be represented by

$$v = v(x) \tag{2-3}$$

Similarly, the v-i curves of an x-controlled current source are independent of the terminal voltage v, and can be represented by

$$i = i(x) \tag{2-4}$$

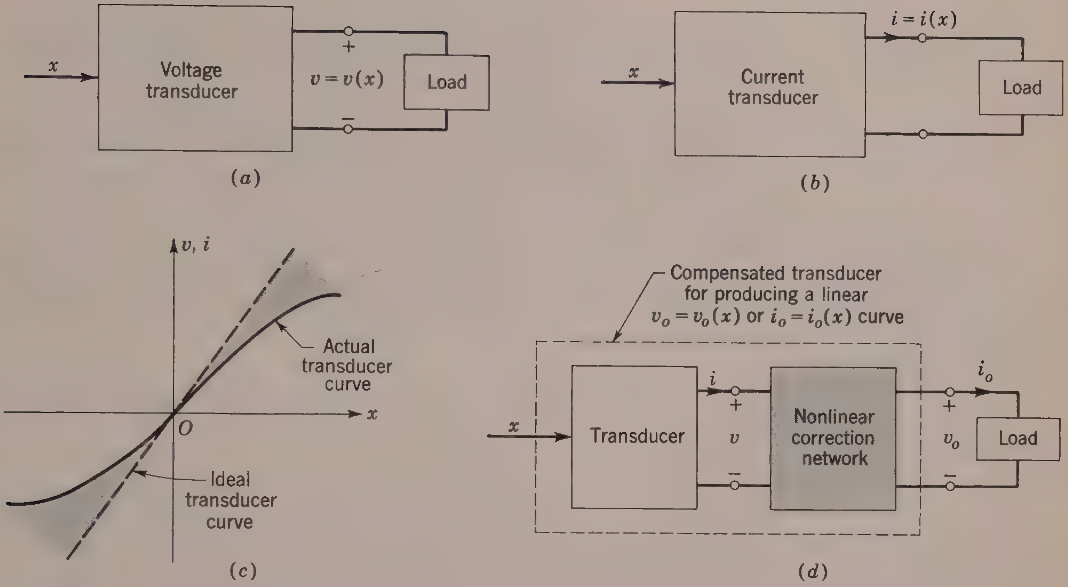


Fig. 2-7. (*a*) A voltage transducer or (*b*) a current transducer is usually characterized by (*c*) a nonlinear v-versus-x or i-versus-x curve. A nonlinear correction network may be designed to linearize this relationship.

Just as there is no such thing as an ideal voltage source or an ideal current source in practice, there is no such thing as an ideal x-controlled voltage source or an ideal x-controlled current source. Nevertheless, we shall find these elements to be extremely important for two reasons. First, the v-i curves of many *practical* controlled resistors consist of near-vertical-line segments or near-horizontal-line segments over some region of the v-i plane. Within this region, the controlled resistor can be considered to be a controlled source. For example, a large region in the first quadrant of a photodiode v-i plane and a large region in the fourth quadrant of a solar cell v-i plane can indeed be considered to be a light-controlled current source. Secondly, a controlled source, even if it does not exist physically, can be combined with other network elements for obtaining a "realistic model" for practical devices. This important application will be demonstrated in Chap. 11.

2-3-1 TRANSDUCERS

When the controlling variable x of a controlled source is a *nonelectrical* variable (e.g., strain, position, pressure, and temperature), the controlled source is usually called a *transducer*.[1] In particular, an x-controlled voltage source is called a *voltage transducer* (Fig. 2-7*a*), and an x-controlled current source is called a *current transducer* (Fig. 2-7*b*). The v-vs.-x or the i-vs.-x relationship of the transducer is generally nonlinear, such as the curve shown in

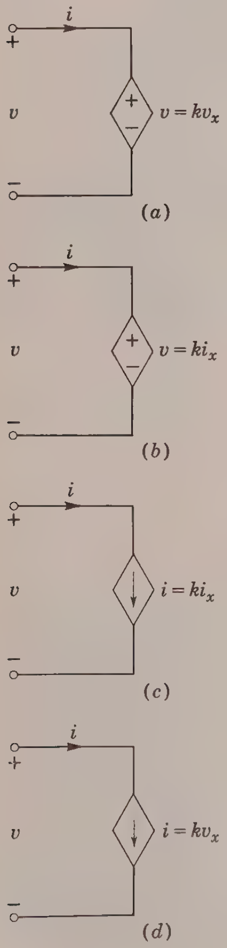
---

[1] The term transducer is sometimes used to denote any controlled element having a nonelectrical controlling variable. In this book we shall use the term transducer only if the device can be modeled by a *controlled voltage source* or a *controlled current source*. Of course, no practical transducer is strictly a voltage source or a current source. However, in Chap. 11 we shall learn methods for realizing a more realistic model.

Fig. 2-7*c*. It is usually desirable to have a linear transducer curve in practice. This can be achieved by designing a nonlinear correction network (Fig. 2-7*d*) by the methods given in Chap. 8.

In practice, transducers are usually built with controlled resistors as building blocks. However, there are some transducers which exist in their original form. For example, a phonograph cartridge is a pressure transducer which transforms the needle pressure into an electrical voltage. For high-fidelity reproduction, the pressure-voltage transducer curve must be as linear as possible. It is important, therefore, that we learn how to design the nonlinear correction network shown in Fig. 2-7*d*.

Fig. 2-8. The symbols for (*a*) a voltage-controlled voltage source; (*b*) a current-controlled voltage source; (*c*) a current-controlled current source; and (*d*) a voltage-controlled current source.



Exercise 1: (*a*) Is a transducer a passive or an active element? If active, explain where the source of energy comes from. (*b*) Why is it desirable to have a linear transducer curve?

Exercise 2: Can you suggest an application for each of the following types of transducers: (*a*) strain-to-voltage transducer; (*b*) position-to-voltage transducer; (*c*) angular-displacement-to-voltage transducer; (*d*) acceleration-to-voltage transducer; (*e*) velocity-to-voltage transducer?

Exercise 3: The relationship of a certain liquid-level-to-voltage transducer (such as a gasoline gage) is given by $v(x) = x^{1/5}$ where x is the liquid level. (*a*) Sketch this relationship, and compare the movement of a voltmeter (connected across the transducer) when the tank is almost full and when the tank is almost empty. (*b*) If a nonlinear correction network is interposed between the voltmeter and the transducer as shown in Fig. 2-7*d*, find the relationship v_o versus v of this network so as to obtain a linear output voltage v_o versus x, namely, $v_o = kx$, where k is a constant.

2-3-2 SOURCES CONTROLLED BY ELECTRICAL VARIABLES

The controlling variable x of an important class of controlled sources is itself an electrical variable, usually a voltage or a current. We shall be mostly concerned in this book with electrically controlled voltage or current sources of the following types: (1) a linear voltage-controlled voltage source whose terminal voltage is linearly related to a controlling voltage v_x, namely, $v = kv_x$, where k is a constant; (2) a linear current-controlled voltage source whose terminal voltage is linearly related to a controlling current i_x, namely, $v = ki_x$; (3) a linear current-controlled current source whose terminal current is linearly related to a controlling current i_x, namely, $i = ki_x$; and (4) a linear voltage-controlled current source whose terminal current is linearly related to a controlling voltage v_x, namely, $i = kv_x$. The symbols for these four types of controlled sources are shown in Fig. 2-8*a* to *d*, respectively.

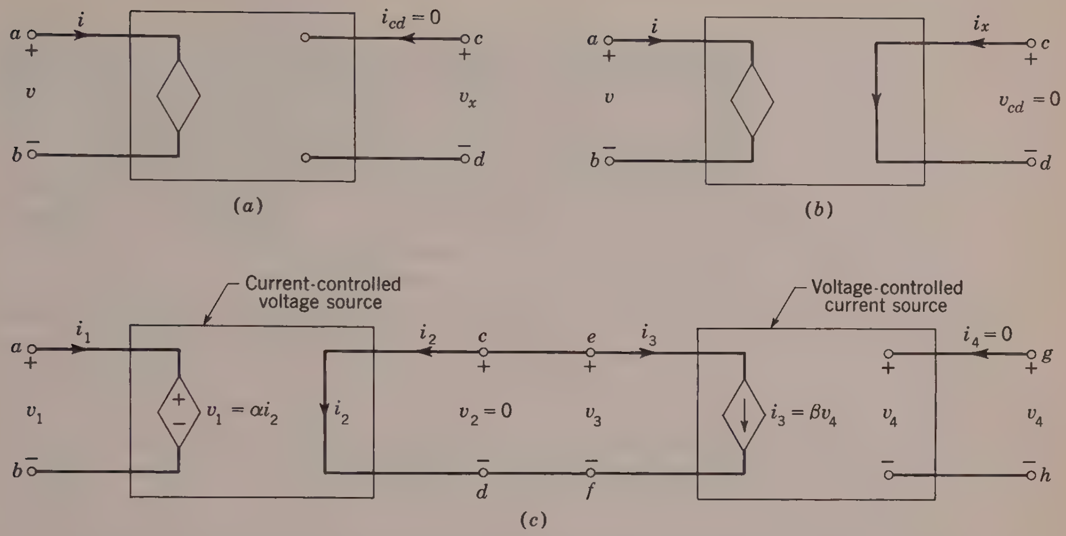


Fig. 2-9. A more explicit symbol of a controlled source is needed when two or more controlled sources are interconnected.

The controlling voltage v_x and the controlling current i_x associated with the above controlled sources may represent the voltage across any terminal pair and the current in any wire, respectively.[1] In the trivial case, v_x and i_x may pertain to a separate network whose values are known. In this case, the controlled source degenerates to an independent source. On the other hand, v_x and i_x may pertain to a separate network, such as a transducer, which in turn depends on a nonelectrical variable x. In many cases, the value of this variable x depends on the output voltage of the original network. In this case, the values of v_x and i_x are not known a priori. We shall even allow a controlled voltage source to depend on its own terminal current, namely,

$$v = ki \tag{2-5}$$

In this case, observe that Eq. (2-5) is simply a relationship between the voltage and current *of the same terminal pair* and is, therefore, equivalent to a linear resistor with a resistance $R = k\Omega$. Similarly, if a controlled current source depends on its own terminal voltage, namely, $i = kv$, then the controlled source is equivalent to a linear resistor with a conductance $G = k$ mhos. We shall find the above interpretation useful in circuit manipulation as well as in modeling.

To avoid confusion, it is sometimes convenient to choose a more explicit symbol for the controlled sources by including the *controlling terminals* as shown in Fig. 2-9*a* for a voltage-controlling variable v_x and in Fig. 2-9*b* for a current-controlling variable i_x.

---

[1] Some authors prefer to restrict v_x and i_x only to those variables pertaining to the same network. We shall not find this restriction necessary in this book.

Observe that $i_{cd} = 0$ in Fig. 2-9*a* because the controlled source is not supposed to draw any current from terminal *c-d* (which may belong to another part of the network). Similarly, $v_{cd} = 0$ in Fig. 2-9*b* because the wire connecting terminal *c-d* is a short circuit (which may represent the wire connected to one terminal of a two-terminal element). The above symbols are preferred in cases where two or more controlled sources are interconnected with one another. For example, consider the circuit shown in Fig. 2-9*c* consisting of a current-controlled voltage source in cascade connection with a voltage-controlled current source. The voltage across terminals *a-b* is

$$v_1 = \alpha\, i_2 \tag{2-6}$$

The current entering terminal *e* is

$$i_3 = \beta\, v_4 \tag{2-7}$$

Applying Kirchhoff current law at terminal *e*, we obtain

$$i_2 = -i_3 \tag{2-8}$$

Substituting Eqs. (2-8) and (2-7) into Eq. (2-6), we obtain

$$v_1 = k\, v_4 \tag{2-9}$$

where $k \equiv -\alpha\beta$. It is important to observe that Eq. (2-9) represents the relationship describing a voltage-controlled voltage source. In other words, we have just proved that a voltage-controlled voltage source can be constructed by connecting a current-controlled voltage source in cascade with a voltage-controlled current source. Similarly, it is easy to show that a current-controlled current source can be constructed by connecting a voltage-controlled current source in cascade with a current-controlled voltage source.

The above four types of linear controlled sources do not exist in their pure forms. However, they can be easily synthesized by appropriate circuits. A practical circuit realization for each is given in Appendix C.

Exercise 1: (*a*) Show that a nonlinear current-controlled voltage source with a terminal voltage $v = v(i_x)$ is equivalent to a nonlinear resistor if the controlling current i_x is equal to the terminal current i. (*b*) Represent the nonlinear resistor *v-i* curve $i = v^3 - 1$ by an appropriate controlled source.

Exercise 2: We could enlarge the class of controlled sources by allowing the controlling variable x to be the time integral, or the time derivatives of some controlling voltage or current. (*a*) Represent a linear capacitor with capacitance C by a controlled source. (*b*) Repeat (*a*) for a linear inductor with inductance L. (*c*) Repeat (*a*) for a nonlinear capacitor with a v-q curve $q = \tanh v$.

Exercise 3: Synthesize a current-controlled current source with terminal current $i = 10\, i_2$ by using a voltage-controlled current source and a current-controlled voltage source.

2-4 A BASIC COMPOSITION TECHNIQUE

As will be shown at various places in this book, certain types of characteristic curves have certain specific applications. With controlled elements, it often occurs that certain characteristic curves fit in very well for a particular design, but unfortunately, the controlling variable x is not what the design calls for. A basic technique for overcoming this difficulty is to find a means to transform the undesirable variable x into the desired variable y. For example, suppose we wish to transform the curve $z = f(x)$ into another curve $z = g(y)$. This can be done if we can find another relation $x = h(y)$ such that

$$z = f(h(y)) \equiv g(y) \tag{2-10}$$

Mathematically, the above procedure of taking the function of another function is known as a *composition* operation. We shall now demonstrate how this technique can be applied.

Consider the cryotron listed in Table 2-1. The controlling variable for this element is the current I_C. Suppose instead of I_C, we would like to use a voltage v_C as the controlling variable. A simple way to accomplish this is to connect a voltage-controlled current source with a terminal current $I_C = k\, v_C$ in place of the independent current source as shown in Fig. 2-10*a*. The constant k can be chosen to be any desired scale factor. In other words, the resulting v-i curves shown in Fig. 2-10*b* are now controlled by the voltage v_C to within a scale factor k.

Another application of the composition technique involves the transformation of a nonelectrical controlling variable into an electrical variable. The most common example is that of transforming light illumination into a controlling voltage. This can be easily accomplished by connecting a voltage-controlled light-emitting source across the element. A highly efficient light-emitting source is the recently discovered light-emitting diode (or gallium arsenide diode). As an example, consider the task of transforming

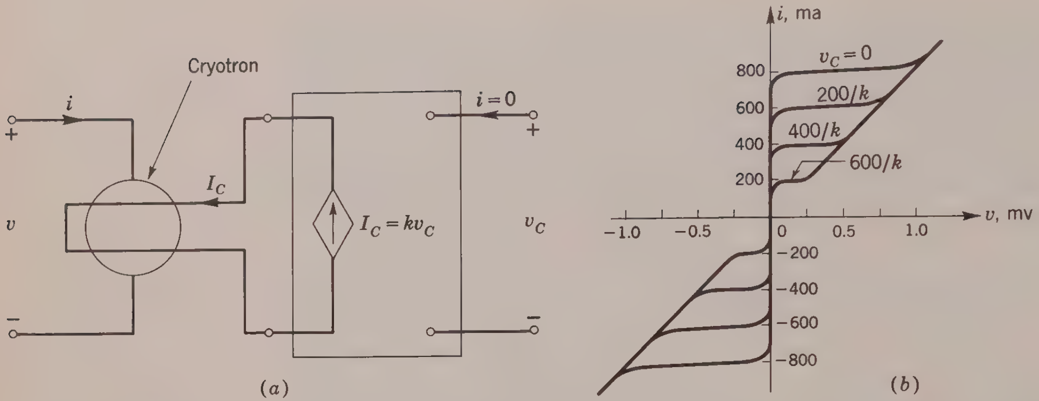
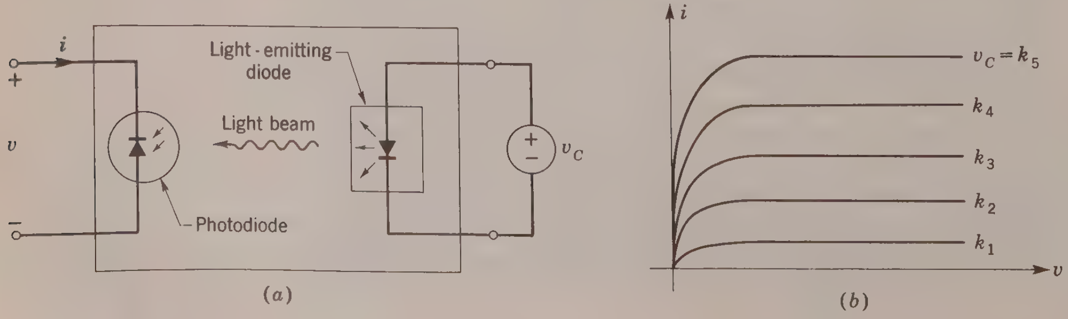


Fig. 2-10. A current-controlled cryotron is transformed into a voltage-controlled cryotron by means of a voltage-controlled current source.

the light-controlled photodiode listed in Table 2-1 into a voltage-controlled diode. This is accomplished by the "coupling" technique shown in Fig. 2-11*a*. The resulting family of *v*-*i* curves shown in Fig. 2-11*b* are now being controlled by the voltage v_C. This technique is widely used in the integrated-circuit industry and is called the *optical-coupling technique.* There are two significant properties which make this technique very attractive. First, the physical dimension of the light-emitting diode and the photodiode can be made extremely small by modern integrated-circuit technology. In fact, it is so small that the user would not even realize that there is a miniature "light bulb" inside the package. Second, the controlling voltage v_C is completely isolated from the photodiode terminals. This means that whereas v_C can affect the *v*-*i* curves, the converse is not true. Hence, there is complete isolation between the input and output terminals.

As a final application of the composition technique, consider the task of transforming a temperature-controlled thermistor into a voltage-controlled resistor. This is commonly done by placing a resistive heating element near the thermistor as shown in Fig. 2-12*a*. The heat produced by the controlling voltage v_C is being transferred to the thermistor body primarily by radiation. This type of volt-

Fig. 2-11. An "optical" coupling technique for transforming a light-controlled photodiode into a voltage-controlled diode.



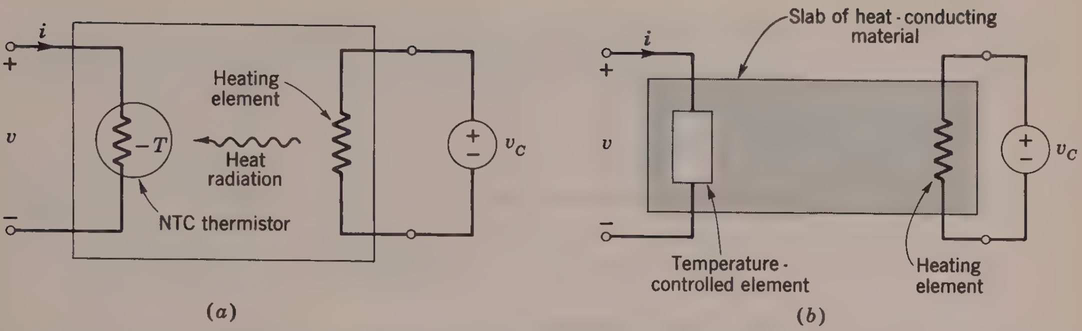


Fig. 2-12. A "thermal" coupling technique for transforming a temperature-controlled element into a voltage-controlled element with delay.

age-controlled thermistor is known commercially as an indirectly heated thermistor. The above technique is often referred to as the *thermal-coupling technique.* It should be observed that an important difference exists between the optical-coupling technique and the thermal-coupling technique. In the former, the coupling can be assumed to take place instantaneously for all practical purposes. In other words, as soon as a voltage pulse is applied, a burst of light will be emitted at the same time. On the other hand, it takes time for the thermal coupling to occur. In other words, if a certain voltage waveform is applied at $t = 0$ sec, it may take a few seconds for the ambient temperature of the thermistor to change to a new value. This behavior makes the thermistor useful only in applications where the time of reaction is not important. There are applications where this "delay" property is taken advantage of. For example, in many control applications it is desirable to have a long delay between the time the controlled voltage is applied and the time when the reaction is to occur. For this type of application, the temperature-controlled element is usually embedded at one end of a slab of heat-conducting material of length l. The heating element is embedded at the other end as shown in Fig. 2-12*b*. Knowing the heat-transfer coefficient of the slab and the desired time delay, the length l can be accurately calculated.

Exercise 1: Suppose the relationship between the illumination F (in footcandles) and the controlling voltage v_C in Fig. 2-11*a* is given by $F = k\,v_C$, and suppose the voltage waveform $v_C(t)$ is given by

$$v_C(t) = \begin{cases} \dfrac{335}{k} \text{ volts} & 0 \leq t < 1\ \mu\text{sec} \\[2ex] \dfrac{3{,}000}{k} \text{ volts} & 1 \leq t < 3\ \mu\text{sec} \\[2ex] \dfrac{1{,}300}{k} \text{ volts} & 3 \leq t < 5\ \mu\text{sec} \end{cases}$$

Find the current $i(t)$ when a 20-volt battery is connected across the photodiode type 1NXX10 given in Appendix D.

Exercise 2: Suppose the heating element of an indirectly heated NTC thermistor produces a heater-voltage-to-temperature relationship $T = -k/\ln(v/100)$, $v < 100$ volts. Find the resistance-vs.-heater-voltage relationship of this thermistor.

2-5 SUMMARY

Two-terminal controlled element A two-terminal network element whose characteristic curve at any time of measurement depends on the value of some physical variable x.

Three types of controlled elements

1. Controlled resistor: characterized by a family of curves $i = i(v,x)$ or $v = v(i,x)$
2. Controlled capacitor: characterized by a family of curves $q = q(v,x)$ or $v = v(q,x)$
3. Controlled inductor: characterized by a family of curves $\varphi = \varphi(i,x)$ or $i = i(\varphi,x)$

Controlled sources A special class of controlled resistors whose family of v-i curves is composed of either vertical lines or horizontal lines.

1. Vertical lines: $v = v(x)$, independent of i. This type is called an x-controlled voltage source.
2. Horizontal lines: $i = i(x)$, independent of v. This type is called an x-controlled current source.

Two classes of controlled sources

1. Transducers: when the controlling variable is a nonelectrical quantity, for example, strain, position, velocity, acceleration.
2. Electrical controlled sources: when the controlling variable is an electrical variable, for example, voltage, current, charge, flux linkage.

Four types of linear electrical controlled sources

1. Voltage-controlled voltage source, $v = k\,v_x$
2. Current-controlled voltage source, $v = k\,i_x$
3. Current-controlled current source, $i = k\,i_x$
4. Voltage-controlled current source, $i = k\,v_x$

Basic composition technique A controlled resistor which is controlled by a nonelectrical variable x can be transformed into an electrically controlled controlled resistor by devising some coupling mechanism for converting the variation in x into a corresponding voltage variation v_C or current variation i_C. Example: given $v = v(i,x)$, choose $x = f(v_C)$ to obtain $v = v(i, f(v_C))$. This procedure for taking the function of another function is called a composition operation.

Two common composition operations

1. Optical coupling
2. Thermal coupling

PROBLEMS

2-1 In the discussion of thermistors, it was shown that if the applied voltage is small (normally less than 5 volts), the thermistor v-i curves are approximately linear. This makes it possible to define a resistance $R(T)$ which depends only on the "ambient" temperature T. However, as the voltage increases, the power dissipation will gradually heat up the thermistor, thereby raising the temperature of the thermistor above its surrounding ambient temperature. This self-heating phenomenon (which may take a minute to reach an equilibrium) is mainly responsible for the nonlinearity of the v-i curves for large voltages. Conversely, for each v-i curve corresponding to an ambient temperature T, we can calibrate the amount of power dissipation at each point on the curve. This information is very useful for designing sensitive "power" measuring devices. In order to obtain this information easily, it is more convenient to plot the family of v-i curves on log-log papers.

(a) Sketch the v-i curves of the NTC thermistor type 1NXX12 on a logarithmic paper.

(b) Draw the constant-power curves $P = vi$ on top of the v-i curves with the parameter $P = 0.01$, 0.1, 1.0, and 10 watts. Verify that these curves represent a system of straight lines and specify the slope of the lines.

(c) Sketch the constant-resistance curve $R = v/i$ on top of the v-i curves with the parameter $R = 0.1$, 0.5, 1.0, 5.0, and 10k Ω. Verify that these curves also represent a system of straight lines and specify the slope of the lines.

(d) Repeat (*b*) and (*c*) on ordinary scales and compare the advantages and disadvantages between the two representations.

2-2 If a current of sufficient size is passed through a thermistor in the nonlinear region of the v-i curve, the thermistor body temperature and consequently its dc resistance $R = v/i$ will depend on the rate of heat transfer from the thermistor to its surrounding. (See Prob. 2-1.) For a fixed amount of power dissipation, a thermistor with an air environment will transfer heat slower than if the environment is

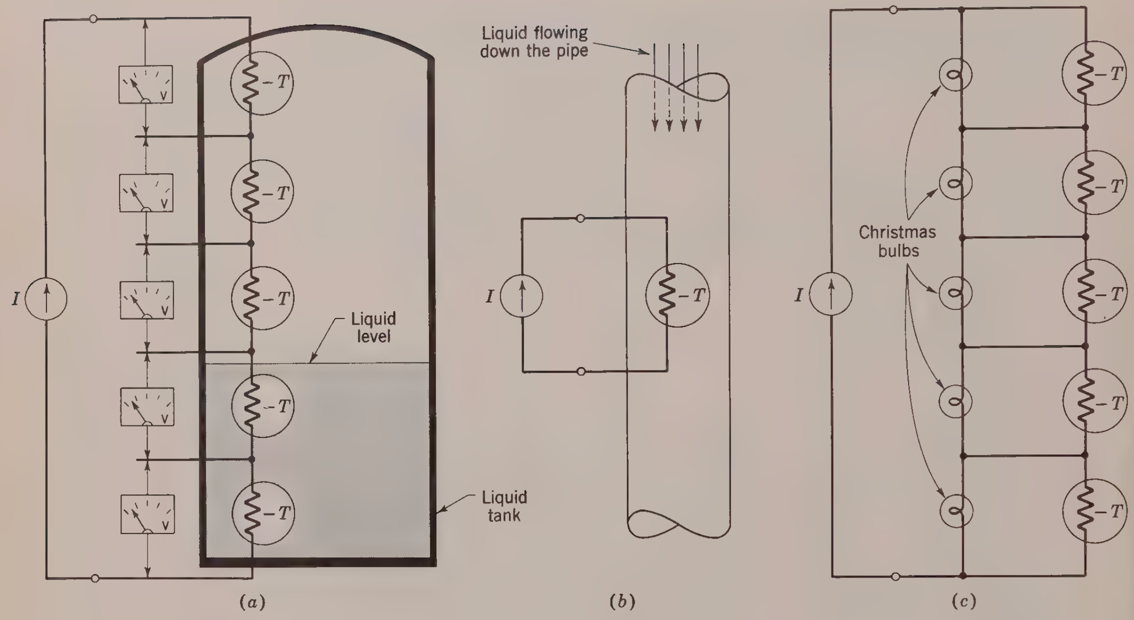


Fig. P2-2.

a liquid (assuming that both air and liquid are at the same ambient temperature). Consequently, the thermistor body temperature will be lower in the liquid than in the air. This property can be used to design liquid-level meters, flowmeters, etc.

(a) Consider the closed tank shown in Fig. P2-2*a* whose wall is lined with a string of identical thermistors type 1NXX12 and connected in series with an $I = 4$ma current source. The voltage developed across each thermistor can be monitored by a voltage-sensitive device such as a voltmeter or even a small light bulb. Assuming that the heat-transfer rate is such that the thermometer equilibrium temperature is 25°F if it is immersed in the liquid and 75°F if it is in air, find the voltage that will be indicated by each voltmeter when the liquid attains a certain level.

(b) The same principle will now be used to design a flowmeter for measuring the velocity of a liquid flowing through a pipe by inserting a thermistor as shown in Fig. P2-2*b*. Explain how to calibrate the scale of a voltmeter to indicate the flow in gallons per second directly? This scale will very likely be nonlinear. What should be done to linearize the scale?

(c) Another application of thermistors is given by the string of Christmas bulbs shown in Fig. P2-2*c*. Without the thermistors, the bulbs will not light whenever one bulb burns out, i.e., becomes an open circuit. With the thermistors connected, the remaining bulbs will continue to light. Explain the principle involved. HINT: If each bulb is lighted, the voltage across it will be small so that the thermistor is not heated, thereby exhibiting a very high resistance.

2-3 Since a solar cell is a low-power active element, it is capable of supplying only a very small amount of power to its external load.

(a) For a given light intensity, there is a point Q on the v-i curve for which maximum power is delivered to the load. Show that this point must satisfy the condition

$$\frac{di}{dv} = -\frac{i}{v}$$

(b) Interpret geometrically the meaning of the above equation and give a procedure for locating the maximum-power point Q.

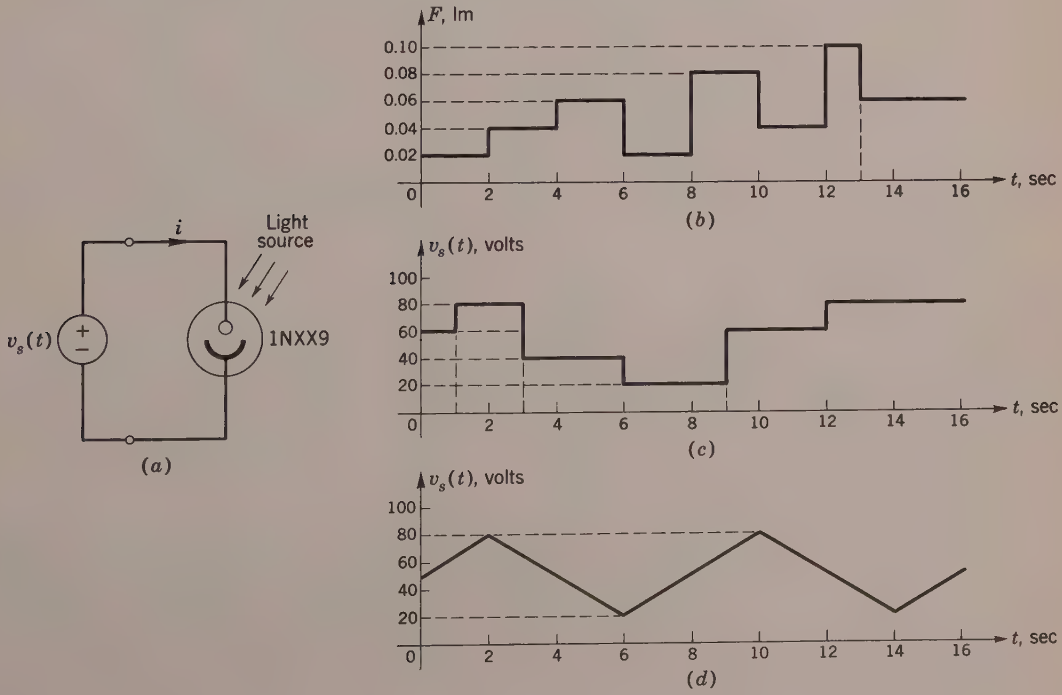
(c) Use the above procedure to find the maximum power that the solar cell type 1NXX11 is capable of supplying with a 350-fc input.

2-4 In a certain biological control system, an artificial light source is used to control the output current of a 1NXX9 gas phototube connected in the simple circuit shown in Fig. P2-4*a*. The light source emits a modulated output luminous flux F as shown in Fig. P2-4*b*.

(a) If the output voltage of the voltage source $v_s(t)$ is as shown in Fig. P2-4*c*, determine the resulting current $i(t)$ through the phototube.

(b) Repeat (*a*) for the case when $v_s(t)$ is as shown in Fig. P2-4*d*.

Fig. P2-4.



(*a*) (*b*) (*c*) (*d*)

2-5 In a certain infrared experiment, a low-frequency square-wave current source $i(t)$ as shown in Fig. P2-5*a* is desired. A simple way to realize this current source consists of focusing a narrow beam of light through the teeth openings of a rotating disk onto a photodiode connected in the circuit shown in Fig. P2-5*b*. Design the complete system so that the current $i(t)$ will be as specified. Use a photodiode type 1NXX10 and a disk with five evenly spaced teeth. Assume that a light source with an illumination of 3,000 fc is available. Find (*a*) the battery voltage E, (*b*) the angles θ_1 (of a space) and θ_2 (of a tooth), and (*c*) the rate of disk rotation in rpm.

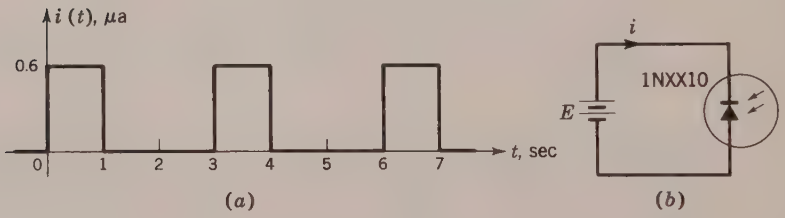


Fig. P2-5.

2-6 Translucency is an important parameter in the manufacture of a certain plastic. It is necessary to monitor this parameter continuously. In order to do this the plastic sheet is passed between a light source and a solar cell type 1NXX11 as shown in Fig. P2-6. When the plastic is properly manufactured, the solar cell receives an illumination of 250 fc. However, the plastic processing varies and the solar cell may receive as much as 500 fc of illumination. These variations must be measured so that the processing may be returned to normal. A voltmeter connected across the solar cell acts as a detector. The voltmeter has an internal resistance of 20 Ω and a maximum rating of 0.25 volt. Select the value of R to be connected in parallel with the voltmeter so that the voltmeter has its maximum deflection when the solar cell receives 500 fc.

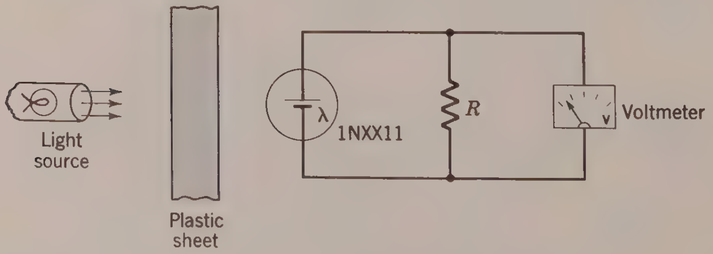
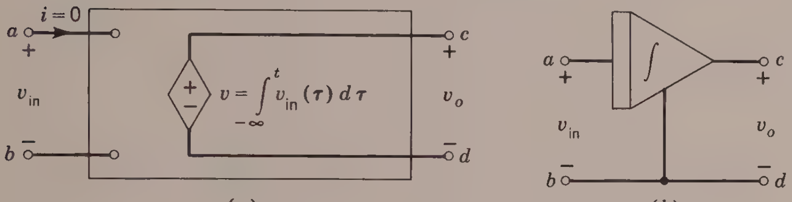


Fig. P2-6.

2-7 The black box shown in Fig. P2-7*a* is called an *integrator* because the output voltage $v_o(t)$ of the controlled voltage source is simply the integral of the input voltage $v_{in}(t)$. For convenience, we usually denote this integrator in practice by the symbol shown in Fig. P2-7*b*.

(a) Using an integrator, devise a simple circuit to display the i-φ curve of a nonlinear inductor on an oscilloscope. You may use a Variac (variable voltage transformer) to provide the desired range of input voltage. You may use a linear resistor to sense the magnitude of current. To avoid grounding problems, you may use a 1:1 isolation transformer.

(b) Repeat (*a*) for the v-q curve of a nonlinear capacitor.



(*a*) (*b*)

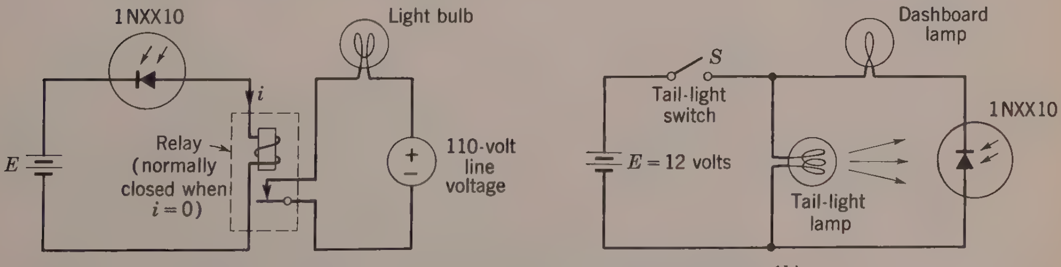
Fig. P2-7.

2-8 A light-controlled resistor can be used for designing many useful gadgets.

(a) A circuit for turning on a house light automatically at dusk is shown in Fig. P2-8*a*. Assuming that the normally closed-relay contact will open as soon as the relay current exceeds 0.05 ma, find the battery voltage E so that the light bulb will turn on as soon as the daylight illumination level drops below 335 fc. You may assume the voltage drop across the relay to be negligible.

(b) An automobile taillight failure warning circuit is shown in Fig. P2-8*b*. The light illumination on the photodiode is derived from the taillight lamp. This circuit, if properly designed, will turn off the dashboard warning lamp whenever the taillight lamp breaks down. If the dashboard lamp requires at least 0.5 ma to operate, how many footcandles of light from the taillight lamp must reach the photodiode?

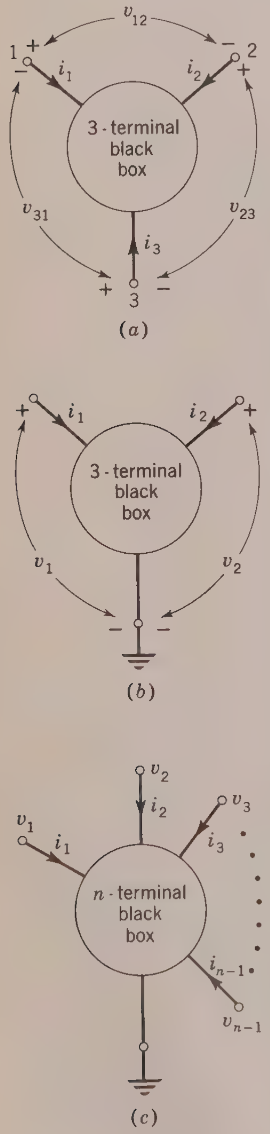
Fig. P2-8.



(*a*) (*b*)

3 MULTITERMINAL ELEMENTS

Fig. 3-1. In the process of characterizing a multiterminal black box, one terminal is arbitrarily chosen to be the common (ground) terminal. The voltages of the remaining terminals are then measured with respect to the common terminal.



3-1 CHARACTERIZATION OF A MULTITERMINAL BLACK BOX

Many devices have more than two terminals. Our object in this chapter is to learn how these multiterminal devices may be characterized so that we shall be in a position to use them effectively. The basic principles discussed in the preceding chapters for characterizing two-terminal devices are still applicable. A set of measurable *independent* variables is selected and a series of external measurements is taken with the objective of deriving a consistent relationship among the variables. Once this relationship is found, we have characterized the black box because from then on, any design using this device can be undertaken on the basis of this relationship alone, thereby obviating the need for further measurements.

To discuss the selection of an independent set of variables, let us consider first the three-terminal black box shown in Fig. 3-1*a*. The most obvious variables are the currents i_1, i_2, and i_3 entering the terminals, and the voltages v_{12}, v_{23}, and v_{31} between terminals. However, as will be shown in the next chapter, the black box in Fig. 3-1*a* can be considered as a *supernode* and the currents i_1, i_2, and i_3 entering this node must satisfy the Kirchhoff current law, namely, $i_1 + i_2 + i_3 = 0$. Hence, if we know the value of any two of these three currents, we can calculate the third, and therefore there is no need to measure all three currents. This observation is equivalent to saying that the three variables i_1, i_2, and i_3 are not independent. Similarly, the three variables v_{12}, v_{23}, and v_{31} are not independent because from the Kirchhoff voltage law, we have $v_{12} + v_{23} + v_{31} = 0$. Consequently, among the six variables shown in Fig. 3-1*a*, only two currents and two voltages are independent. For this reason, we may select any terminal to be the common terminal (usually called the *ground*) and define the two currents i_1, i_2 and the voltages v_1, v_2 as shown in Fig. 3-1*b*. In theory, there is no reason for preferring one terminal over another as the common terminal. In practice, however, such a preference may be desirable because the measurements may be easier and more accurately obtained.[1] To avoid ambiguity, it is of the utmost importance to specify the common terminal associated with the measured characteristics of a particular device.

With the above precaution, let us investigate the type of measurements that may be taken. Just as in the two-terminal case, it is necessary to excite the black box by a voltage source or a current source. However, the response to these excitations need not be restricted to currents and voltages. Recall that it is possible to measure the charge q_j entering terminal j by integrating the current i_j, namely,

$$q_j(t) = \int_{-\infty}^{t} i_j(\tau)\, d\tau \qquad j = 1, 2 \tag{3-1}$$

Similarly, we can measure the flux linkage φ_j associated with each voltage v_j between terminal j and the ground by integrating the voltage v_j, namely,

$$\varphi_j(t) = \int_{-\infty}^{t} v_j(\tau)\, d\tau \qquad j = 1, 2 \tag{3-2}$$

Hence, among the variables of interest to us are v_j, i_j, q_j, and φ_j, where $j = 1$ or 2. Any *independent* combination of these variables constitutes a valid set of measurements.[2] If a certain set of measurements leads to some consistent relationship, then the device is said to be characterized by that relationship. Among the several distinct combinations, the following three cases are of special practical importance:

Case 1: Relationship involving only v_1, v_2; i_1, i_2

Case 2: Relationship involving only v_1, v_2; q_1, q_2

Case 3: Relationship involving only i_1, i_2; φ_1, φ_2

In analogy to the classification of two-terminal elements, any three-terminal element which falls under case 1, case 2, or case 3 is called a *three-terminal resistor,* a *three-terminal capacitor,* or a *three-terminal inductor,* respectively.

It is now a simple matter to extend our definition to multiterminal elements ($n > 3$). For the same reason as before, among the n currents entering the terminals and the n voltages between terminals, only $(n-1)$ currents and $(n-1)$ voltages are independent. Hence, one of the terminals is chosen to be common, and the $(n-1)$ independent terminal currents and voltages are defined as shown in Fig. 3-1*c*. The corresponding element classifications now take the following forms:

---

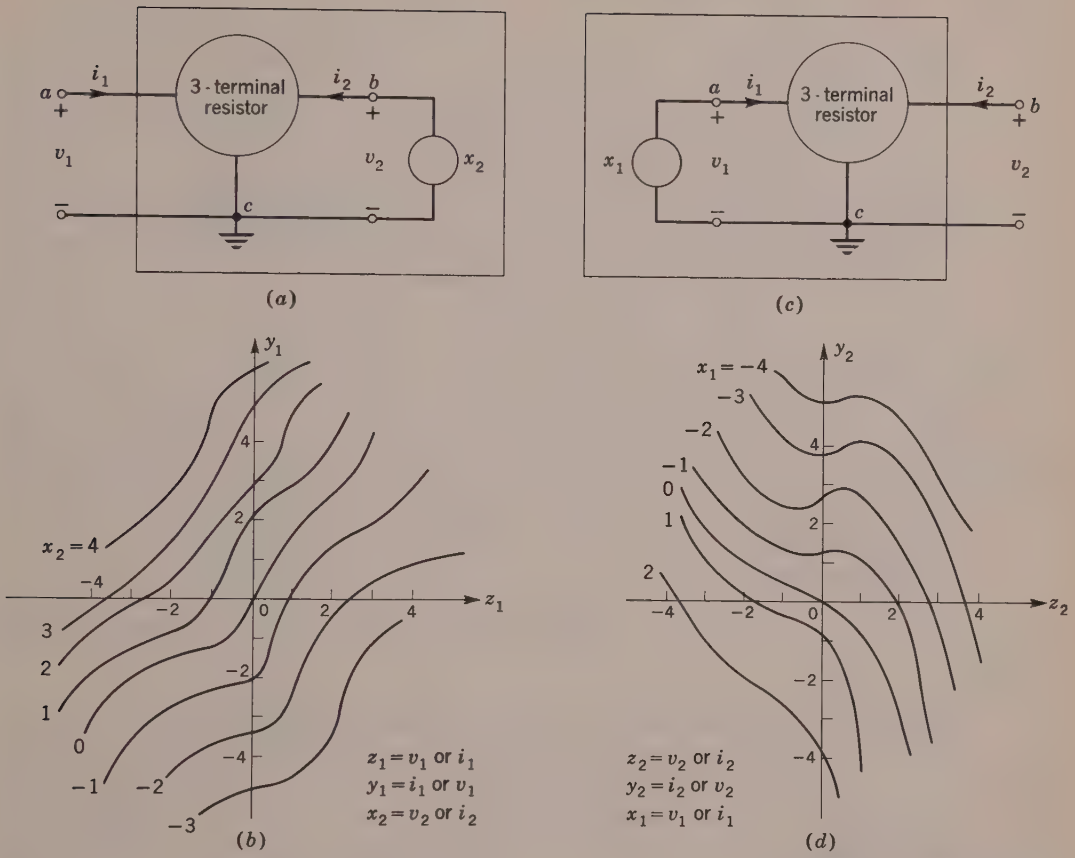
[1] This is especially true for transistors, where the characteristic curves can be more accurately, and more easily, measured if a particular terminal, called the emitter, is chosen to be the ground terminal. This terminal is usually identified by the transistor manufacturer in the form of appropriate markings.

[2] Observe that the combination i_j and q_j is not valid because these variables are already related by Eq. (3-1) and are therefore not independent. Similarly, the combination v_j and φ_j is not valid.

1. N-terminal resistors, involving only $v_1, v_2, \ldots, v_{n-1}; i_1, i_2, \ldots, i_{n-1}$
2. N-terminal capacitors, involving only $v_1, v_2, \ldots, v_{n-1}; q_1, q_2, \ldots, q_{n-1}$
3. N-terminal inductors, involving only $i_1, i_2, \ldots, i_{n-1}; \varphi_1, \varphi_2, \ldots, \varphi_{n-1}$

3-2 THREE-TERMINAL RESISTORS

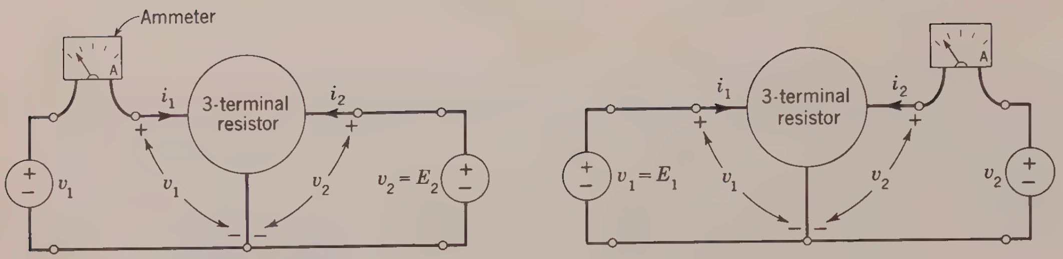
A three-terminal resistor is characterized by two separate sets of measurements. One set of measurements is obtained by applying an independent source x_2 (which may be a voltage source or a current source) across terminals *b-c* as shown in Fig. 3-2*a*, and measuring the v_1-i_1 curve across terminals *a-c* of the resulting two-terminal black box. For each value of the independent source x_2, a corresponding v-i curve is obtained. If enough measurements are taken, a whole family of v_1-i_1 curves can be plotted, each curve corresponding to a given value of x_2. For convenience, we shall usually refer to this set of measurements as the *input-characteristic curves.* An example of a hypothetical set of input-characteristic curves is shown in Fig. 3-2*b*, where the variables z_1-y_1 may denote either v_1-i_1 or i_1-v_1. Observe that the input-characteristic curves can be interpreted as the v-i curves of a controlled resistor with x_2 as the controlling variable. As a matter of fact, there are some practical applications for which the output terminals *b-c* are connected across an independent source. In this case, the three-terminal resistor is used simply as a controlled resistor, and the set of input characteristic curves is sufficient for all design purposes. However, if the circuit connected across the output terminals *b-c* is not an independent source, which is usually the case, then the set of input characteristic curves does not completely characterize the three-terminal resistor. This is because we have no way of finding the value of x_2 which applies at a particular time. To complete the characterization, we need to perform another set of measurements.



(a) (b) (c) (d)

Fig. 3-2. A three-terminal resistor is characterized by two sets of curves; each set can be interpreted as the v-i curves of a two-terminal controlled resistor.

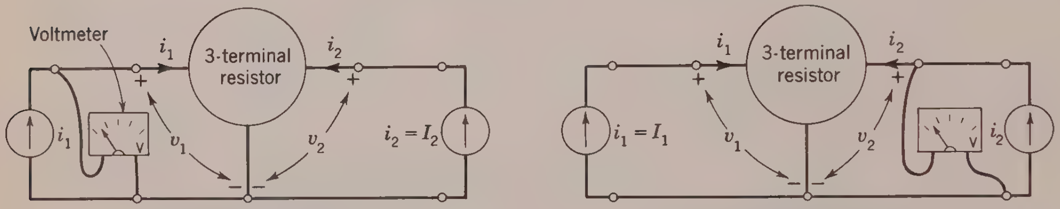
The second set of measurements is obtained by applying another independent source x_1 (which can be either a voltage source or a current source) across terminals a-c as shown in Fig. 3-2c, and measuring the v_2-i_2 curve across terminals b-c of the resulting two-terminal black box. An analogous set of v_2-i_2 curves can be plotted with the value of x_1 as a parameter. We shall usually refer to this set of curves as the *output-characteristic curves.* A hypothetical set of output-characteristic curves is shown in Fig. 3-2d, where the variables z_2-y_2 may denote either v_2-i_2 or i_2-v_2. Observe that the output-characteristic curves of a three-terminal resistor can also be interpreted as the v-i curves of a controlled resistor. In this sense, a three-terminal resistor is completely equivalent to two controlled resistors.



(a) Independent variable: v_1
Controlling variable: v_2
$i_1 = i_1(v_1, v_2)$

Method 1

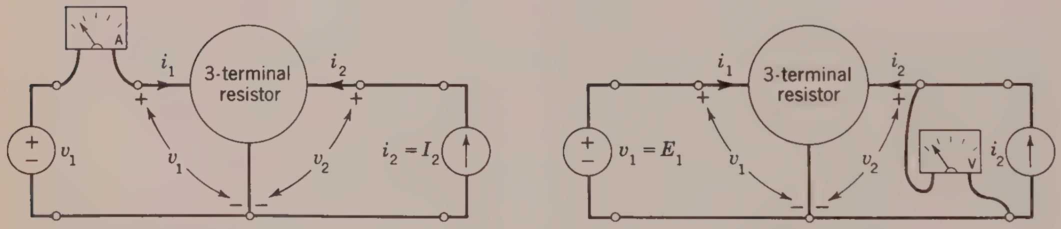
(b) Independent variable: v_2
Controlling variable: v_1
$i_2 = i_2(v_2, v_1)$



(a) Independent variable: i_1
Controlling variable: i_2
$v_1 = v_1(i_1, i_2)$

Method 2

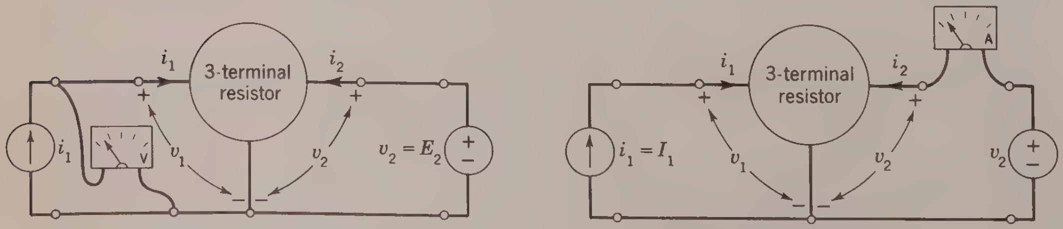
(b) Independent variable: i_2
Controlling variable: i_1
$v_2 = v_2(i_2, i_1)$



(a) Independent variable: v_1
Controlling variable: i_2
$i_1 = i_1(v_1, i_2)$

Method 3

(b) Independent variable: i_2
Controlling variable: v_1
$v_2 = v_2(i_2, v_1)$



(a) Independent variable: i_1
Controlling variable: v_2
$v_1 = v_1(i_1, v_2)$

Method 4

(b) Independent variable: v_2
Controlling variable: i_1
$i_2 = i_2(v_2, i_1)$

Fig. 3-3. Four laboratory setups for measuring the characteristic curves of a three-terminal resistor.

3-2-1 FORMS OF REPRESENTATION

In Fig. 3-2*b*, y_1 represents the dependent variable, z_1 the independent variable, and x_2 the controlling variable. Similarly, in Fig. 3-2*d*, y_2, z_2, and x_1 represent the dependent, independent, and controlling variables, respectively. Since each of these variables may be either a voltage or a current, there are four distinct combinations corresponding to the four laboratory setups shown in Fig. 3-3. In the first setup, the input-characteristic curves are obtained by measuring the current i_1 corresponding to different values of v_1, while keeping the value v_2 fixed at some value E_2. The resulting points are plotted into a v_1-i_1 curve, with $v_2 = E_2$ as the parameter. The procedure is repeated for different values of v_2. To obtain the output-characteristic curves, we measure the current i_2 corresponding to different values of v_2, while keeping the value of v_1 fixed at some value E_1. The resulting points are plotted into a v_2-i_2 curve, with $v_1 = E_1$ as the parameter.

Exercise: Explain how the measurements are taken in the remaining three setups and sketch a hypothetical set of input- and output-characteristic curves in each case.

In order to illustrate the different combinations, let us consider a hypothetical three-terminal resistor.[1] Suppose the characteristic curves of this resistor have been measured by method 1 and are shown in Fig. 3-4*a*. In order to simplify our discussion, we have "contrived" these curves so that they may be represented mathematically by the following two equations:[2]

$$i_1 = (v_1 + v_2)^3 \tag{3-3}$$

$$i_2 = (v_2 - v_1)^{1/3} \tag{3-4}$$

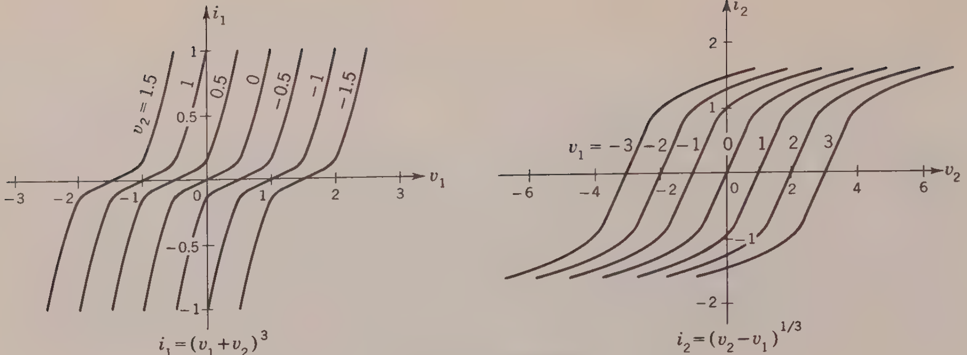
Since the slope at each point of the curves has the unit of "conductance," we shall refer to these sets of curves as the *conductance representation.* It is important to realize that each of the four methods of measurements given in Fig. 3-3 gives a complete characterization of the device. Therefore, for a given device, it is necessary to choose only one method, whichever is simpler, and from the resulting measurements we should be able to replot the points so as to obtain two new sets of curves corresponding to another method of measurement.

In order to illustrate how one set of measurements can be converted into an equivalent set of measurements, let us repeat our

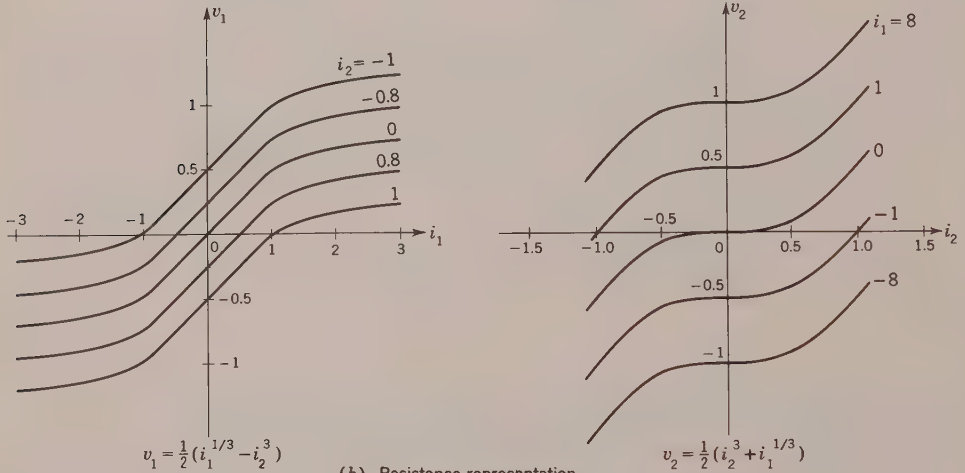
---

[1] The reader is strongly urged to think in terms of hypothetical devices, even though they may not be as motivating as real devices. After all, the transistor was just as hypothetical as many of our examples until its discovery in 1948! The effort to think more abstractly will be amply rewarded by the ability to analyze circuits containing new devices yet to be invented.

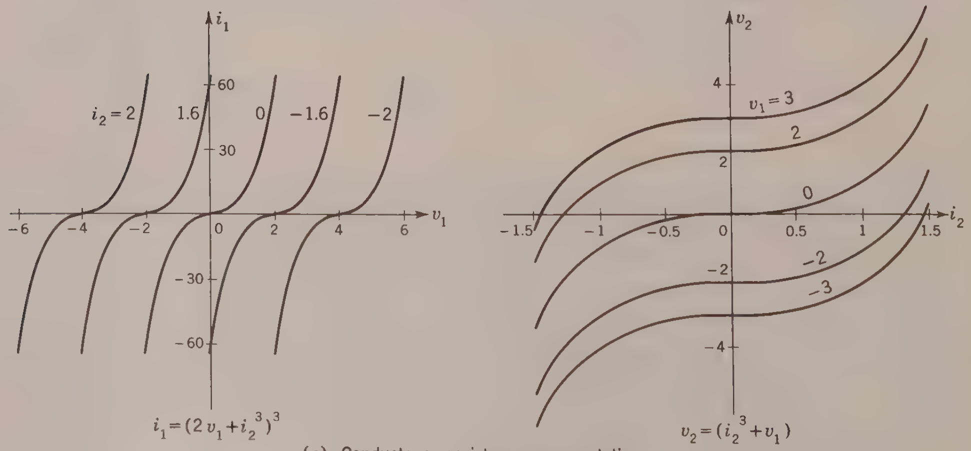
[2] In practice, it is generally not possible to find a *reasonably* simple mathematical description. This is why the majority of practical three-terminal devices is invariably described by two sets of characteristic curves.



(*a*) Conductance representation

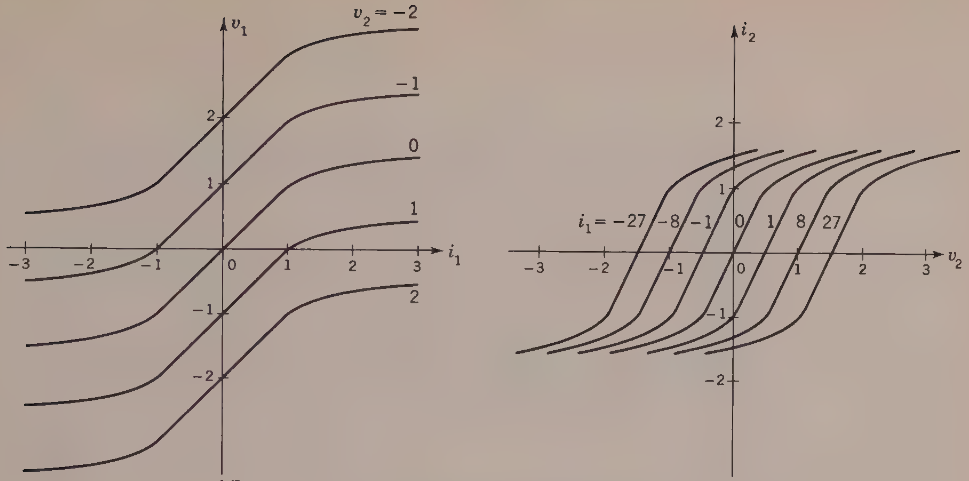


(*b*) Resistance representation



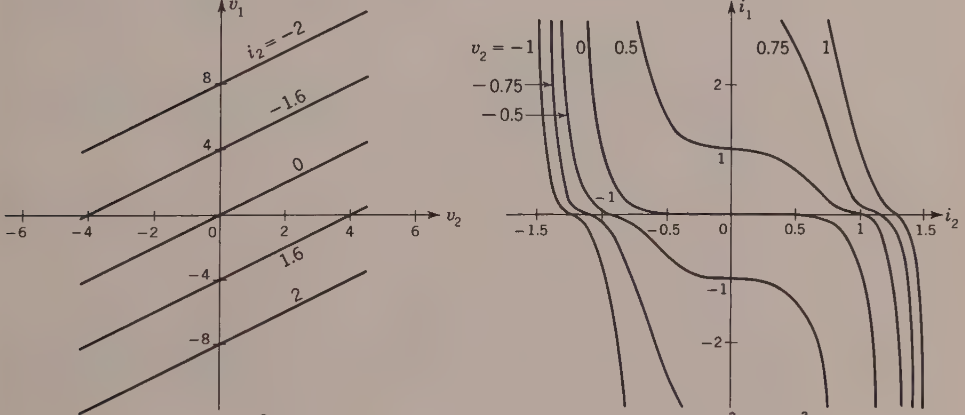
(*c*) Conductance-resistance representation

Fig. 3-4. A three-terminal resistor is completely characterized by any one of six distinct forms of representations.



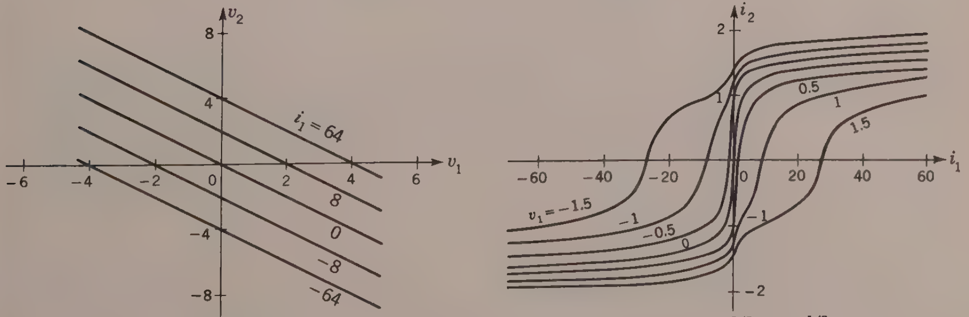
$v_1 = (i_1^{1/3} - v_2)$ $i_2 = (2v_2 - i_1^{1/3})^{1/3}$

(*d*) Resistance-conductance representation



$v_1 = (v_2 - i_2^3)$ $i_1 = (-i_2^3 + 2v_2)^3$

(*e*) Voltage ratio - current ratio representation



$v_2 = (-v_1 + i_1^{1/3})$ $i_2 = (i_1^{1/3} - 2v_1)^{1/3}$

(*f*) Current ratio - voltage ratio representation

Fig. 3-4 *(Continued)*

experiment on the above device using the remaining three methods. If these experiments were carried out accurately on this device, we should obtain the sets of curves shown in Fig. 3-4*b*, *c*, and *d*, respectively. For analogous reasons, these characteristic curves are called the *resistance representation, conductance-resistance representation,* and *resistance-conductance representation,* respectively. To show that we can derive the resistance representation from the conductance representation, let us solve for v_1 and v_2 in terms of i_1 and i_2 in Eqs. (3-3) and (3-4). The result is given by

$$v_1 = \tfrac{1}{2}(i_1^{1/3} - i_2^3) \tag{3-5}$$

$$v_2 = \tfrac{1}{2}(i_2^3 + i_1^{1/3}) \tag{3-6}$$

Equations (3-5) and (3-6), when plotted, must give the same sets of curves shown in Fig. 3-4*b*, because they pertain to the same device.

By a similar procedure, we can solve for i_1 and v_2 in terms of v_1 and i_2 from Eqs. (3-3) and (3-4) to obtain the following conductance-resistance representation:

$$i_1 = (2v_1 + i_2^3)^3 \tag{3-7}$$

$$v_2 = (i_2^3 + v_1) \tag{3-8}$$

Similarly, by solving for v_1 and i_2 in terms of i_1 and v_2 from Eqs. (3-3) and (3-4), we obtain the following resistance-conductance representation:

$$v_1 = (i_1^{1/3} - v_2) \tag{3-9}$$

$$i_2 = (2v_2 - i_1^{1/3})^{1/3} \tag{3-10}$$

Actually, we can solve for two more combinations from Eqs. (3-3) and (3-4) which do not have an experimental counterpart. For the first combination, we can solve for v_1 and i_1 in terms of v_2 and i_2 to obtain

$$v_1 = (v_2 - i_2^3) \tag{3-11}$$

$$i_1 = (-i_2^3 + 2v_2)^3 \tag{3-12}$$

Since the slope of each point on the first set of curves represents a voltage ratio and that of the second set represents a current ratio, we shall refer to the above sets of relationships as the *voltage ratio–*

current ratio representation. The second combination is obtained by solving for v_2 and i_2 in terms of v_1 and i_1; thus

$$v_2 = (-v_1 + i_1^{1/3}) \tag{3-13}$$

$$i_2 = (i_1^{1/3} - 2v_1)^{1/3} \tag{3-14}$$

In order to avoid ambiguity with the voltage ratio—current ratio representation, we shall refer to this combination as the *current ratio–voltage ratio representation.* Observe that the last two representations cannot be measured directly; they must be converted from the other four representations.

Only six common forms of representation are given in Fig. 3-4. Actually, there are other forms of representation which are just as valid and, in fact, may be more appropriate for some applications. In particular, any two sets of measurements which are *independent* of one another would qualify. For that matter, any representation on the left column of Fig. 3-4 may be combined with any representation on the right column. For example, it is just as valid for us to represent the three-terminal resistor by Eq. (3-7) for its input-characteristic curves and Eq. (3-4) for its output-characteristic curves.

The above examples show conclusively that whereas a three-terminal resistor may be characterized by at least six distinct forms of representations, these representations are not independent. If one representation is given analytically, then the remaining representations may be obtained by solving for the desired dependent variables in terms of the independent variables, provided such a solution exists. Even in the two-terminal case, we have seen many examples of v-i curves described by $v = v(i)$ for which it is not possible to solve for $i = i(v)$. This is because the latter may be a multivalued function of v. The same situation holds for three-terminal resistors. We may not be able to solve for v_1 and v_2 in terms of i_1 and i_2 in the resistance representation if the variables v_1 and v_2 are multivalued functions of i_1 and i_2. For example, consider the representation

$$i_1 = (v_1 + v_2)^2 \tag{3-15}$$

$$i_2 = v_1^2 + v_2^4 \tag{3-16}$$

It is easy to see that there are many values of (v_1,v_2) which give the same (i_1,i_2). For example, $v_1 = 1$, $v_2 = 1$ and $v_1 = -1$, $v_2 = -1$ both give $i_1 = 4$ and $i_2 = 2$.

There is a special type of three-terminal resistor which will be of some importance to us when the characteristic curves consist of a system of uniformly spaced straight lines. For this type of three-terminal resistor it is not necessary to plot the characteristic curves because they can be easily described by two linear equations. Corresponding to the six distinct representations given in Fig. 3-4, we now have the following simplified equations:

Conductance representation

$$\begin{aligned} i_1 &= g_{11}v_1 + g_{12}v_2 + I_1 \\ i_2 &= g_{21}v_1 + g_{22}v_2 + I_2 \end{aligned}$$

Resistance representation

$$\begin{aligned} v_1 &= r_{11}i_1 + r_{12}i_2 + E_1 \\ v_2 &= r_{21}i_1 + r_{22}i_2 + E_2 \end{aligned}$$

Conductance-resistance representation

$$\begin{aligned} i_1 &= k_{11}v_1 + k_{12}i_2 + I_1 \\ v_2 &= k_{21}v_1 + k_{22}i_2 + E_2 \end{aligned}$$

Resistance-conductance representation

$$\begin{aligned} v_1 &= h_{11}i_1 + h_{12}v_2 + E_1 \\ i_2 &= h_{21}i_1 + h_{22}v_2 + I_2 \end{aligned}$$

Voltage ratio–current ratio representation

$$\begin{aligned} v_1 &= t_{11}v_2 + t_{12}i_2 + E_1 \\ i_1 &= t_{21}v_2 + t_{22}i_2 + I_1 \end{aligned}$$

Current ratio–voltage ratio representation

$$\begin{aligned} v_2 &= \gamma_{11}v_1 + \gamma_{12}i_1 + E_2 \\ i_2 &= \gamma_{21}v_1 + \gamma_{22}i_1 + I_2 \end{aligned}$$

(3-17)

A further special case occurs when the constants I_1, I_2, E_1, and E_2 in Eqs. (3-17) are zero. In this case, each of six sets of coefficients g_{ij}, r_{ij}, k_{ij}, h_{ij}, t_{ij}, or γ_{ij} completely describes the element. These coefficients can be easily calculated from the system of straight lines. For example, the g_{ij} parameters can be found from Eq. (3-17); thus

$$\begin{aligned} g_{11} &= \left.\frac{i_1}{v_1}\right|_{v_2=0} & \qquad g_{12} &= \left.\frac{i_1}{v_2}\right|_{v_1=0} \\ g_{21} &= \left.\frac{i_2}{v_1}\right|_{v_2=0} & \qquad g_{22} &= \left.\frac{i_2}{v_2}\right|_{v_1=0} \end{aligned} \tag{3-18}$$

Equation (3-18) says that to find g_{11}, we look for the straight line with parameter $v_2 = 0$ and calculate its slope. Physically, referring to Fig. 3-3*a*, g_{11} can be found by short-circuiting the output terminals and measuring the conductance across the input terminals. The remaining three parameters have a similar interpretation. For this reason, the g_{ij} parameters are usually called *short-circuit conductances*. For a similar reason, the r_{ij} parameters are usually called *open-circuit resistances;* the k_{ij} and h_{ij} parameters are usually called *hybrid parameters;* the t_{ij} and γ_{ij} parameters are usually called *transmission parameters.*

Exercise 1: Derive Eqs. (3-5) to (3-14) from Eqs. (3-3) and (3-4).

Exercise 2: A three-terminal resistor is characterized by the following voltage ratio–current ratio representation:

$v_1 = (v_2 + i_2)^3$

$i_1 = (v_2 + i_2)^{1/3}$

(*a*) Find the remaining five representations. (*b*) Do all representations consist of a family of curves? If some representation reduces to only one curve, explain its significance.

Exercise 3: The resistance representation of a certain three-terminal resistor is given by

$v_1 = 2i_1 - 3i_2 + 2$

$v_2 = i_1 + \frac{1}{2} i_2 + 4$

(*a*) Find the remaining five representations. (*b*) Sketch the characteristic curves in each of the six representations.

Exercise 4: (*a*) Find the conditions on the characteristic curves so that the constants I_1, I_2, E_1, and E_2 are zero in Eqs. (3-17). (*b*) Give explicit formulas for evaluating the remaining parameters r_{ij}, k_{ij}, h_{ij}, t_{ij}, and γ_{ij} in a form similar to Eq. (3-18). (*c*) Using these formulas, calculate the numerical value in each case for the resistor given in Exercise 3, and check your answers with those obtained in Exercise 3.

Exercise 5: Repeat Exercise 3 for the following representation:

$v_1 = i_1 + 2i_2$

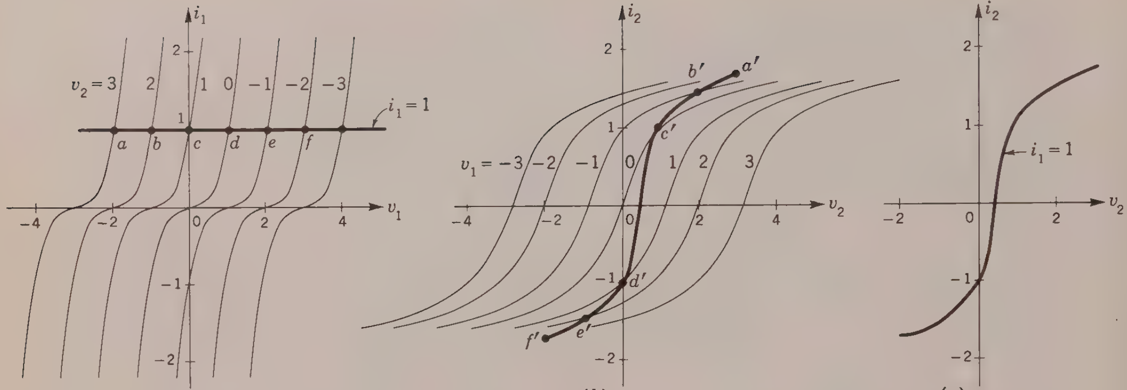
$v_2 = 2i_1 + 4i_2$

If any representation fails to exist, give a geometrical interpretation and explain why.

3-2-2 GRAPHICAL TRANSFORMATION OF REPRESENTATIONS

The fact that a certain representation may not exist analytically does not imply that this representation cannot be obtained by other means. On the contrary, we can always use graphical techniques to replot a given set of characteristic curves into another set. As a matter of fact, this approach is more realistic because almost all practical devices are described by two sets of curves, not by two simple equations.

To illustrate the graphical approach, let us suppose it is desirable to convert the curves given in the conductance representation (Fig. 3-4*a*) into an equivalent set of curves in the resistance-conductance representation of Fig. 3-4*d*. Consider first the determination of the input-characteristic curves $v_1 = v_1(i_1,v_2)$.



(a) (b) (c)

Fig. 3-5. The graphical procedure for obtaining the output-characteristic curves of the resistance-conductance representation consists of (*a*) draw the horizontal line $i_1 = k$ in the v_1-i_1 plane of the conductance representation; (*b*) transfer the points of intersection into the v_2-i_2 plane; and (*c*) replot this curve with i_1 as the parameter.

Since v_2 is the controlling variable, we must assign specific parameter values to v_2 and try to locate the points corresponding to this parameter value. These points can be found systematically by observing that the constraint $v_2 = k$ is satisfied by all points on the v-i curve corresponding to $v_2 = k$ in Fig. 3-4*a*. Hence, in this case, we need only to replot this curve with i_1 as the horizontal axis and v_1 the vertical axis.[1] Repeating this procedure for different values of k would lead to the same characteristic curves shown in Fig. 3-4*d*.

To obtain the output-characteristic curves, we observe that the controlling parameter is now i_1. Hence, for each parameter value k the constraint $i_1 = k$ represents a horizontal line in the input-characteristic curves of Fig. 3-4*a*, as shown in Fig. 3-5*a*. The coordinates (v_1,v_2) of the points of intersection (such as *a*, *b*, *c*, *d*, *e*, and *f*) of this line with the input-characteristic curves can now be transferred to corresponding points in the v_2-i_2 plane as shown in Fig. 3-5*b* (points a', b', c', d', e', and f'). If a smooth curve is passed through these points, we would obtain the desired v_2-i_2 curve corresponding to the controlling parameter $i_1 = k$. This curve is replotted as shown in Fig. 3-5*c*. If the procedure is repeated for different parameter values, we would obtain the same output-characteristic curves shown in Fig. 3-4*d*.

Exercise 1: Give the graphical procedure for obtaining the curves in the resistance representation from the curves in the conductance representation. Sketch one curve in each case and check your answer with that given in Fig. 3-4*b*.

Exercise 2: Repeat Exercise 1 for the following cases: (*a*) conductance-resistance representation and (*b*) current ratio–voltage ratio representation.

Exercise 3: Given the characteristic curves in the resistance-conductance representation, find the graphical methods for obtaining the remaining five representations.

---

[1] A simple method for obtaining these curves is to trace the curves of Fig. 3-4*a* on transparent paper, flip the paper over, and reposition it so that the v_1 axis becomes the vertical axis.

3-2-3 TRANSFORMATION TO ANOTHER COMMON TERMINAL

Our discussions and derivations in the preceding sections have been presented under the tacit assumption that terminal *c* of the three-terminal resistor in Fig. 3-2 is the ground terminal. In many practical problems, occasions arise wherein a different ground terminal is desired. We can certainly go to the laboratory and perform the same sets of measurements as before with respect to the new ground terminal. However, not only is this a time-consuming process, but on many occasions we do not have the device on hand. For example, we may be using the characteristic curves supplied by the manufacturer, and we certainly would not want to buy the device and measure the curves (with respect to our new ground terminal) only to find out later that the resulting characteristic curves are unsuitable. Since the same device is involved, there must be a way to transform the characteristic curves (with respect to our common terminal) into an equivalent set of characteristic curves with respect to another terminal. Indeed we can; in fact, all we need to do is to apply Kirchhoff voltage and current laws.

To show the procedure for obtaining the desired transformation, let us consider the same three-terminal resistor of Fig. 3-2. For comparison purposes, this resistor is redrawn in Fig. 3-6*a* with its original voltage and current variables. The same resistor is shown in Fig. 3-6*b*, but with a new set of voltage and current variables with respect to the new ground terminal *b*. In order to avoid confusion, we use primes to denote the new variables. The original resistor is characterized by Eqs. (3-3) and (3-4), which we reproduce here for convenience:

$$i_1 = (v_1 + v_2)^3 \tag{3-19}$$

$$i_2 = (v_2 - v_1)^{1/3} \tag{3-20}$$

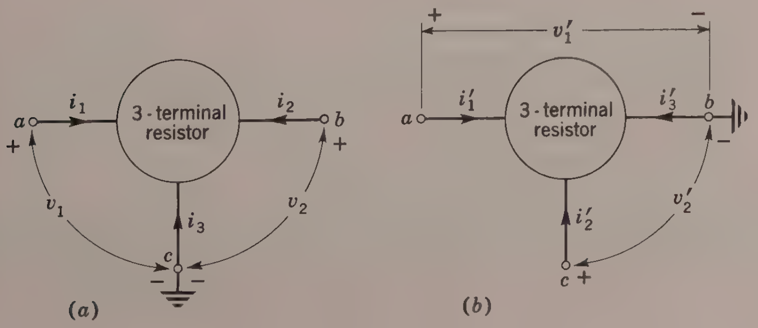


Fig. 3-6. Corresponding to each ground terminal, a distinct set of voltage and current variables must be defined for a three-terminal resistor.

In order to transform to the new variables, we must express v_1, v_2, i_1, and i_2 in terms of v_1', v_2', i_1', and i_2'. Since the two resistors shown in Fig. 3-6 are identical, their corresponding voltages and currents must be identical. Hence, we can apply the Kirchhoff voltage law to obtain

$$v_1 = v_1' - v_2' \tag{3-21}$$

$$v_2 = -v_2' \tag{3-22}$$

Similarly, since the Kirchhoff current law is applicable to the "supernode" discussed earlier, $i_1 + i_2 + i_3 = 0$. But $i_1 = i_1'$ and $i_3 = i_2'$; therefore, we obtain

$$i_1 = i_1' \tag{3-23}$$

$$i_2 = -(i_1' + i_2') \tag{3-24}$$

Substituting Eqs. (3-21) to (3-24) into Eqs. (3-19) to (3-20), we obtain, after some simplification, the following desired relationship:

$$i_1' = (v_1' - 2v_2')^3 \tag{3-25}$$

$$i_2' = (v_1')^{1/3} - (v_1' - 2v_2')^3 \tag{3-26}$$

In most practical devices, the curves are given graphically. In this case, the same method applies; however, the procedure is much more tedious. To do this systematically, a table must be prepared for values computed from Eqs. (3-21) to (3-24). This task can be automated by writing a computer program for converting one representation into another.

Exercise 1: Find the procedure for transforming the conductance representation of the three-terminal resistor shown in Fig. 3-6*a* into an equivalent conductance representation with terminal *a* as the ground terminal. Let i_1'' and i_2'' denote the currents entering terminals *b* and *c*, respectively. Let v_1'' and v_2'' denote the voltage at terminals *b* and *c*, respectively.

Exercise 2: By a strictly graphical approach, transform the characteristic curves of Fig. 3-4*a* into an equivalent set of curves in the conductance representation but with terminal *b* as the ground terminal. Check your solution by plotting the curves given by Eqs. (3-25) and (3-26).

3-2-4 SOME PRACTICAL THREE-TERMINAL RESISTORS

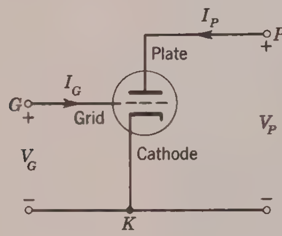
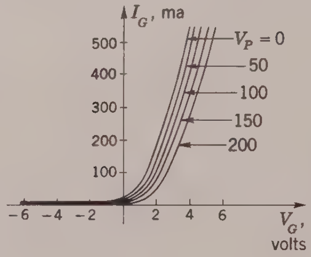
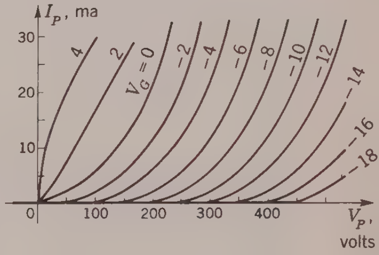
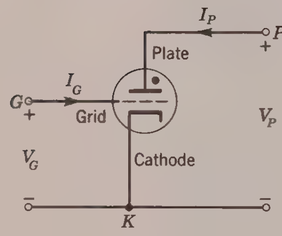
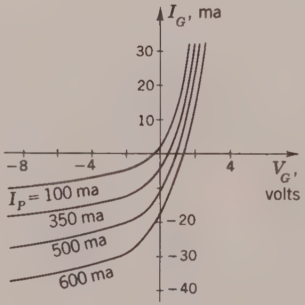
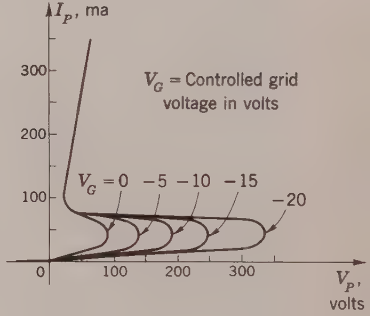
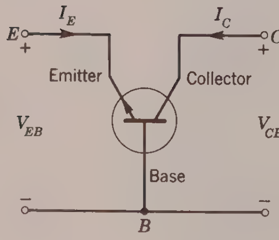
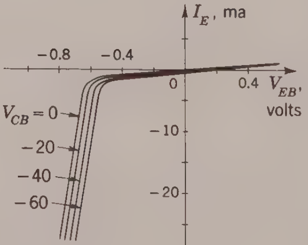
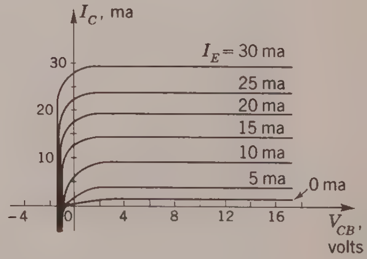
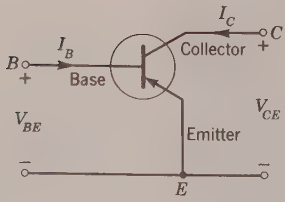
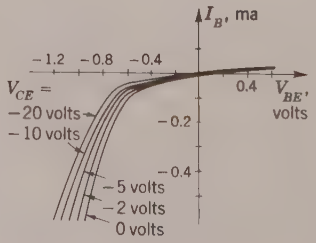
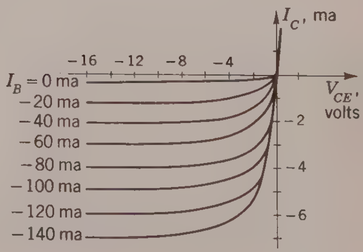
So far, for the sake of generality, we have considered only hypothetical elements. Actually, there are many practical three-terminal resistors which are widely available commercially, some of which are listed in Table 3-1. Because they are practical devices, the three terminals of each device are usually identified by terms related to internal construction. For example, for the *vacuum triode,* these terminals are called the *grid,* the *plate,* and the *cathode.* Instead of using subscripts 1 and 2, we shall use letter subscripts for the terminal current and voltage variables. The characteristic curves of the triode are given in the conductance representation. Observe that whereas the output-characteristic curves are widely spaced, the input-characteristic curves are "packed" close to each other. Physically, this means that whereas the grid voltage V_G exerts a significant influence on the output-characteristic curves, the plate voltage V_P influences the input-characteristic curves only very slightly. In view of this property, we can usually approximate the input-characteristic curves by only one curve for most practical purposes. Mathematically, this is equivalent to replacing the representation $i_1 = i_1(v_1,v_2)$ by $i_1 = i_1(v_1)$. As we shall see later, this approximation greatly simplifies the analysis procedure.

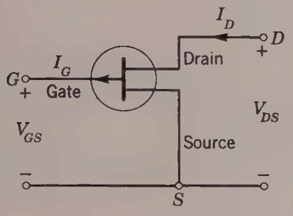
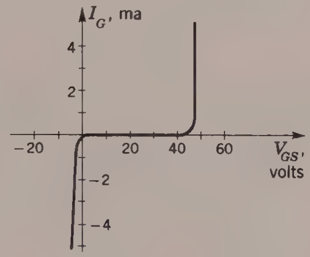
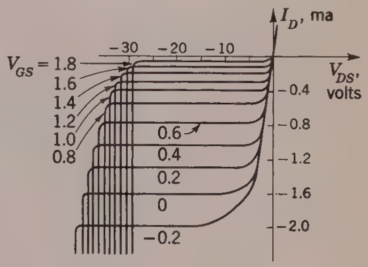
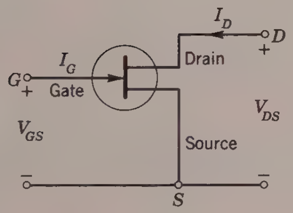
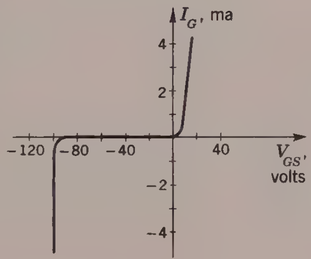
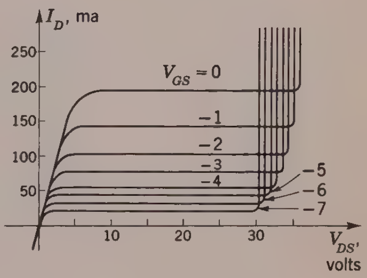
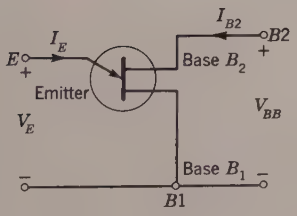
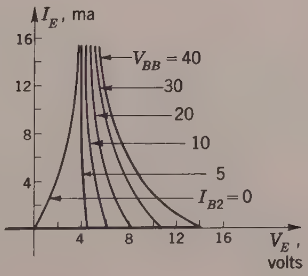
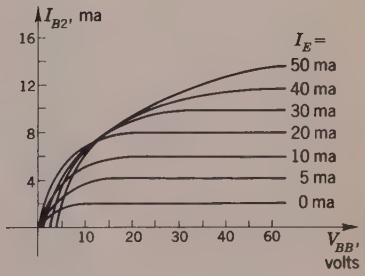
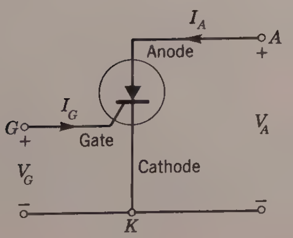
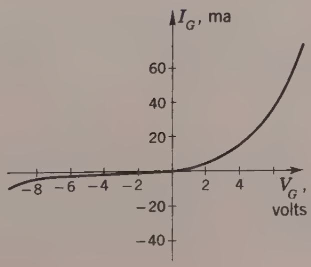
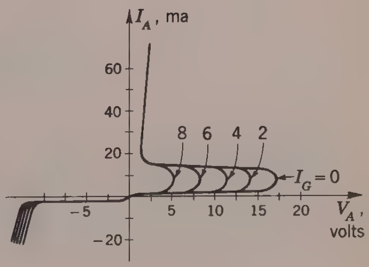
The second three-terminal resistor listed is called a *thyratron.* Observe that the input-characteristic curves are given in the conductance-resistance representation, but the output-characteristic curves are given in the conductance representation. This particular representation is usually chosen by the manufacturer because the measurements can be more accurately taken. The next two elements listed are called *n-p-n* and *p-n-p transistors,* respectively. In both cases, the terminals are called *emitter, base,* and *collector.* The only difference in their symbols is the direction of the arrowhead, *leaving* the emitter terminal for the *n-p-n* transistor and *entering* the emitter terminal for the *p-n-p* transistor. In both cases, the input-characteristic curves are given in the conductance representation, but the output-characteristic curves are given in the resistance-conductance representation. The characteristic curves for the *n-p-n* transistor in Table 3-1 are measured with the base as the common terminal. Hence, they are called the *common-base characteristic curves.* In contrast with this, the emitter of the *p-n-p* transistor in Table 3-1 is the ground terminal. Consequently, these characteristic curves are called the *common-emitter characteristic curves.* In view of our previous discussion, it does not matter to which configuration the characteristic curves belong; after all, given one, we can find the other if necessary.[1] However, it turns out that for either the *n-p-n* or the *p-n-p* transistors, the common-emitter characteristic curves are easier to measure and are, therefore, the favorite form supplied by manufacturers. Just

---

[1] See Appendix D for more *n-p-n* and *p-n-p* transistor characteristic curves in both configurations.

TABLE 3-1 *A list of common three-terminal resistors.*

<table>
<tr><th>Name</th><th>Symbol</th><th>Input-characteristic curves</th><th>Output-characteristic curves</th></tr>
<tr><td>Vacuum triode</td><td>
</td><td>
</td><td>
</td></tr>
<tr><td>Thyratron</td><td>
</td><td>
</td><td>
</td></tr>
<tr><td><i>n-p-n</i> transistor</td><td>
</td><td>
</td><td>
</td></tr>
<tr><td><i>p-n-p</i> transistor</td><td>
</td><td>
</td><td>
</td></tr>
</table>

Name	Symbol	Input-characteristic curves	Output-characteristic curves
P-channel field-effect transistor (FET)	 I_D D + Drain I_G G + Gate V_{GS} V_{DS} Source – – S		
N-channel field-effect transistor (FET)	 I_D D + Drain I_G G + Gate V_{GS} V_{DS} Source – – S		
Unijunction transistor	 I_{B2} B2 + I_E E + Base B_2 Emitter V_E V_{BB} Base B_1 – – B1		
Silicon-controlled rectifier (SCR)	 I_A A + Anode I_G G + Gate V_G V_A Cathode – – K		

as with the vacuum triode, the input characteristics of either the *n-p-n* or *p-n-p* transistors are usually approximated by a single *v-i* curve.

The next two elements listed in Table 3-1 are called *p-channel and n-channel field-effect transistors*, respectively.[1] The three terminals are called the *gate*, the *drain*, and the *source*. The usual characteristic curves supplied by the manufacturer are in the common-source configuration, as shown in Table 3-1. Observe that unlike the preceding cases, the input-characteristic curves of the FET consist of only one *v-i* curve. Physically, this means that the drain voltage and the drain current have no effect whatsoever on the input-characteristic curves. As we shall see later, this property is highly desirable for many applications.

The next element in the list is called a *unijunction transistor*. The three terminals are usually called the *emitter*, the *base* B_1, and the *base* B_2. Unlike the other elements, both the input and the output-characteristic curves are widely spread out. It is not possible, therefore, to approximate the input-characteristic curves by only one *v-i* curve.

The last element in the list is called a *silicon-controlled rectifier* (SCR). The three terminals are usually called the *gate*, the *anode*, and the *cathode*. For most practical purposes, the input-characteristic curves consist of a single *v-i* curve. Observe also that above a certain value of anode current, the output-characteristic curves reduce to an almost vertical line.

What are three-terminal resistors good for? They are capable of doing whatever a two-terminal resistor can do, and a lot more that a two-terminal resistor cannot do. This follows from the observation that any three-terminal resistor can be used as a two-terminal controlled resistor by connecting an independent voltage or current source across two terminals of the three-terminal resistor. For example, as we shall see later, the silicon-controlled rectifier and the thyratron are normally used in this manner. In some cases, we may simply short-circuit or open-circuit the two terminals. For example, if we short-circuit the plate and the cathode terminals of the vacuum triode, the resulting input-characteristic curve is simply the V_G-I_G curve corresponding to $V_P = 0$. But this is just the *v-i* curve of the vacuum diode shown in Table 1-1. Similarly, we can obtain a first-quadrant *v-i* curve similar to that of the constant-current diode shown in Table 1-1 by short-circuiting the gate and source terminals of an *n*-channel FET. Hence, whereas a two-terminal resistor may be considered as a special case of a controlled resistor, a controlled resistor may be considered as a special case of a three-terminal resistor. Because of their greater generality, three-terminal resistors are more versatile and flexible. Consequently, they are the most widely used in practice.

---

[1] This element is usually called the FET in practice. It is fast becoming a workhorse in electronic circuits.

Exercise 1: Show how to connect a transistor to obtain a two-terminal resistor with a *v-i* curve similar to that of a junction diode. Do this for the following cases: (*a*) *n-p-n* transistor type 2NXX15; (*b*) *p-n-p* transistor type 2NXX16; (*c*) *n-p-n* transistor type 2NXX17A; (*d*) *n-p-n* transistor type 2NXX17B; (*e*) *p-n-p* transistor type 2NXX18A; and (*f*) *p-n-p* transistor type 2NXX18B.

Exercise 2: Find the current leaving the base B_1 of the unijunction transistor type 2NXX21 when a 20-volt battery is connected across terminals B_2 and B_1, with terminal E left unconnected.

Exercise 3: Find the voltage from the "drain" to the "gate" and the current leaving the "source" of a *p*-channel FET type 2NXX20. Assume that a 5-volt battery is connected between the drain and the source (positive polarity), and assume that the gate is connected to the source.

Exercise 4: A three-terminal element is said to be *passive* if the sum of the instantaneous powers $p(t) \equiv v_1 i_1 + v_2 i_2$ is never negative. Otherwise, it is said to be *active*. (*a*) What restrictions must we impose on the locations of the characteristic curves in order for a three-terminal resistor to be passive? (*b*) Classify each of the resistors listed in Table 3-1 according to whether it is active or passive.

3-3 THREE-TERMINAL CAPACITORS

A three-terminal element was previously defined in Sec. 3-1 to be a three-terminal capacitor if it can be characterized by two sets of curves, or relationships, involving the variables v_1, v_2, q_1, and q_2. The same procedure for measuring three-terminal resistor characteristic curves is applicable here. However, instead of measuring the terminal currents i_1 and i_2, we measure their time integrals $q_1(t) = \int_{-\infty}^{t} i_1(\tau)\,d\tau$ and $q_2(t) = \int_{-\infty}^{t} i_2(\tau)\,d\tau$. This can be achieved with the integrator to be described in Sec. 3-6-2.

Just as for three-terminal resistors, there are several possible forms of representation. Since the principles are identical, only one form will be discussed here, namely,

$$q_1 = q_1(v_1,v_2) \tag{3-27}$$

$$q_2 = q_2(v_1,v_2) \tag{3-28}$$

To find the currents $i_1(t)$ and $i_2(t)$ corresponding to any voltage waveforms $v_1(t)$ and $v_2(t)$, we apply the chain rule for differentiation to Eqs. (3-27) and (3-28); thus

$$i_1(t) = \frac{dq_1}{dt} = \frac{\partial q_1}{\partial v_1}\frac{dv_1(t)}{dt} + \frac{\partial q_1}{\partial v_2}\frac{dv_2(t)}{dt} \tag{3-29}$$

$$i_2(t) = \frac{dq_2}{dt} = \frac{\partial q_2}{\partial v_1}\frac{dv_1(t)}{dt} + \frac{\partial q_2}{\partial v_2}\frac{dv_2(t)}{dt} \tag{3-30}$$

In the special case where Eqs. (3-27) and (3-28) are linear functions of v_1 and v_2, namely,

$$q_1(v_1,v_2) = C_{11}v_1 + C_{12}v_2 \tag{3-31}$$

$$q_2(v_1,v_2) = C_{21}v_1 + C_{22}v_2 \tag{3-32}$$

the capacitor is said to be a *linear capacitor,* and Eqs. (3-29) and (3-30) simplify to[1]

$$i_1(t) = C_{11}\frac{dv_1(t)}{dt} + C_{12}\frac{dv_2(t)}{dt} \tag{3-33}$$

$$i_2(t) = C_{21}\frac{dv_1(t)}{dt} + C_{22}\frac{dv_2(t)}{dt} \tag{3-34}$$

Although three-terminal capacitors are not yet commercially available, the possibility of their future availability cannot be dismissed. As a matter of fact, the parasitic capacitances between the three terminals of any practical element can be considered as forming a three-terminal linear capacitor. Although undesirable, this parasitic capacitor must be included in the analysis if its associated coefficients C_{11}, C_{12}, C_{21}, and C_{22} are significant. This situation usually occurs in the design of integrated circuits.

Exercise 1: Devise a laboratory setup for measuring the characteristic curves of a three-terminal capacitor.

Exercise 2: Discuss all possible forms for representing the characteristic curves of a three-terminal capacitor.

3-4 THREE-TERMINAL INDUCTORS

A three-terminal element is called a *three-terminal inductor* if it can be characterized by two sets of curves, or relationships, involving the variables i_1, i_2, φ_1, and φ_2. The measurement procedure is analogous to that of the three-terminal resistor. However, instead of measuring the voltages v_1 and v_2, we measure their time integrals $\varphi_1(t) = \int_{-\infty}^{t} v_1(\tau)\,d\tau$ and $\varphi_2(t) = \int_{-\infty}^{t} v_2(\tau)\,d\tau$.

---

[1] In analogy with the capacitance of a two-terminal linear capacitor, the coefficients C_{11} and C_{22} are usually called *self-capacitances,* whereas the coefficients C_{12} and C_{21} are called *mutual capacitances.*

Among the various possible forms of representation, consider the following more common case:

$$\varphi_1 = \varphi_1(i_1,i_2) \tag{3-35}$$

$$\varphi_2 = \varphi_2(i_1,i_2) \tag{3-36}$$

To find the voltages $v_1(t)$ and $v_2(t)$ corresponding to any current waveforms $i_1(t)$ and $i_2(t)$, we apply the chain rule, thereby obtaining

$$v_1(t) = \frac{d\varphi_1}{dt} = \frac{\partial\varphi_1}{\partial i_1}\frac{di_1(t)}{dt} + \frac{\partial\varphi_1}{\partial i_2}\frac{di_2(t)}{dt} \tag{3-37}$$

$$v_2(t) = \frac{d\varphi_2}{dt} = \frac{\partial\varphi_2}{\partial i_1}\frac{di_1(t)}{dt} + \frac{\partial\varphi_2}{\partial i_2}\frac{di_2(t)}{dt} \tag{3-38}$$

In the special case of a three-terminal linear inductor, Eqs. (3-35) and (3-36) reduce to[1]

$$\varphi_1 = L_{11}i_1 + L_{12}i_2 \tag{3-39}$$

$$\varphi_2 = L_{21}i_1 + L_{22}i_2 \tag{3-40}$$

In this case, the inductor terminal voltages simplify to

$$v_1(t) = L_{11}\frac{di_1(t)}{dt} + L_{12}\frac{di_2(t)}{dt} \tag{3-41}$$

$$v_2(t) = L_{21}\frac{di_1(t)}{dt} + L_{22}\frac{di_2(t)}{dt} \tag{3-42}$$

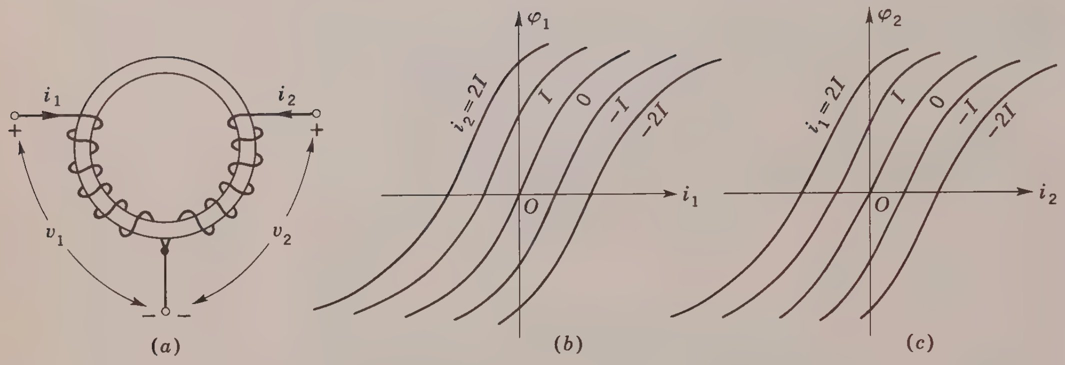
The most common example of a commercially available three-terminal inductor is that of a toroidal coil with a center tap, as shown in Fig. 3-7*a*. The corresponding sets of *i*-φ curves are shown in Fig. 3-7*b* and *c*.

Exercise 1: Devise a laboratory setup for measuring the characteristic curves of a three-terminal inductor.

Exercise 2: Discuss all possible forms for representing the characteristic curves of a three-terminal inductor.

---

[1] In analogy with the inductance of a two-terminal linear inductor, the coefficients L_{11} and L_{22} are usually called *self-inductances,* whereas the coefficients L_{12} and L_{21} are called *mutual inductances.*

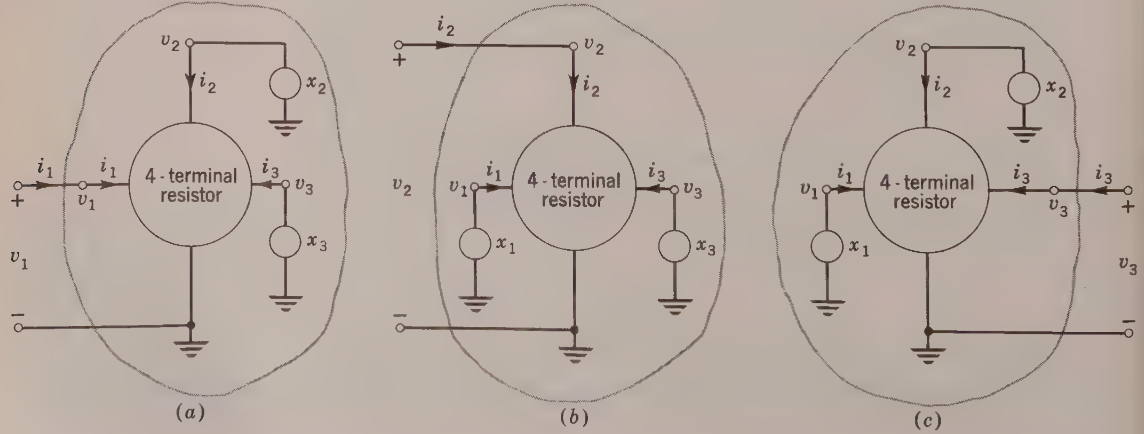


(*a*) (*b*) (*c*)

Fig. 3-7. A common three-terminal inductor characterized by two sets of parallel-spaced i-φ curves.

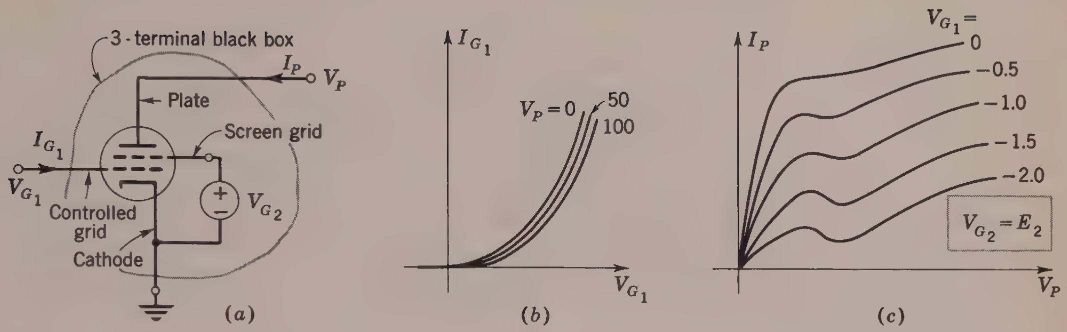
3-5 MULTITERMINAL ELEMENTS

Consider now an n-terminal black box with $n > 3$. Conceptually, the procedure for characterizing this black box is similar to the three-terminal case. For example, three distinct laboratory setups (Fig. 3-8) are necessary to *completely* characterize a four-terminal resistor. In each setup, the black box is converted into a two-terminal controlled resistor with two controlling variables. Since it is not possible to plot points in a four-dimensional space, it is necessary to fix the value of one controlling variable and measure a set of v-i curves with the second controlling variable as a parameter. Therefore, many separate sets of curves are needed for each of the three setups. More generally, $n - 1$ distinct laboratory setups are required to completely characterize an n-terminal resistor. Each setup alone would involve as many sets of v-i curves



(*a*) (*b*) (*c*)

Fig. 3-8. To characterize a four-terminal black box, three distinct laboratory setups are necessary; each setup involves as many sets of measurements as are necessary to include all desired combinations of parameter values of the controlling variables.



(a) (b) (c)

Fig. 3-9. A tetrode, when operating as a three-terminal resistor, can be characterized by two sets of characteristic curves by fixing the value of the screen grid voltage V_{G_2}.

as are necessary to include all desired combinations of parameter values of the controlling variables. It should be clear by now that it is in general impractical to completely characterize an *n*-terminal black box. However, we may partially characterize an *n*-terminal resistor by operating the resistor as a three-terminal controlled resistor.

Exercise: Draw the laboratory setups necessary to completely characterize a five-terminal resistor. Explain how the measurements are to be taken.

3-5-1 MULTITERMINAL RESISTORS OPERATING AS THREE-TERMINAL CONTROLLED RESISTORS

Any *n*-terminal resistor can be converted into a three-terminal controlled resistor by connecting a voltage source or a current source from each of the remaining *n*-3 terminals to the common terminal. Once the terminal voltage of the voltage sources and the terminal currents of the current sources are fixed, the resulting three-terminal black box is no different from an ordinary three-terminal resistor,[1] and can therefore be characterized by the procedures described in Sec. 3-4. The only precaution that we must observe is that any circuit containing an *n*-terminal resistor (operating as a three-terminal controlled resistor) which is designed in accordance with a set of characteristic curves (usually furnished by the manufacturer) must be operated in practice with the same controlling voltages and currents for which these characteristic curves are measured.

A commercially available four-terminal resistor which is usually operated as a three-terminal controlled resistor is the *tetrode* shown in Fig. 3-9*a*. The controlling terminal is called the *screen grid* and the remaining three terminals are called the *controlled grid* (no. 1), the *plate,* and the *cathode,* respectively. The characteristic curves corresponding to a typical screen grid voltage V_{G_2} are shown in Fig. 3-9*b* and *c*.

---

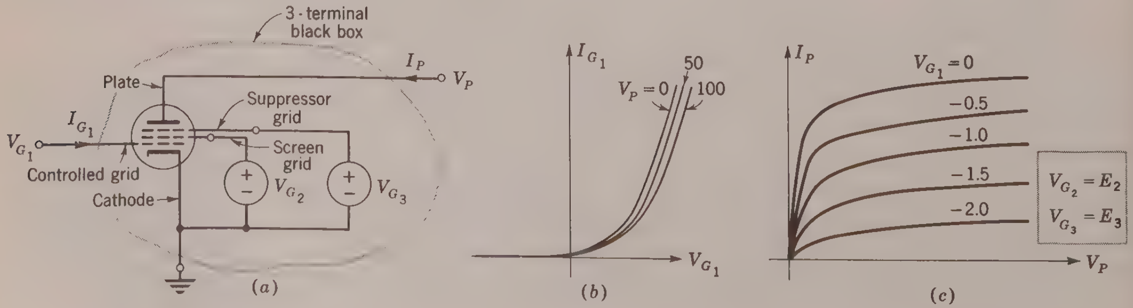
[1] If the terminal voltage of a voltage source, or the terminal current of a current source, is set equal to zero, the corresponding terminal is simply short-circuited, or open-circuited, with respect to the ground terminal.

A commercially available five-terminal resistor which is usually operated as a three-terminal controlled resistor is the *pentode* shown in Fig. 3-10*a*. The two controlling terminals are called the *screen grid* (no. 2) and *suppressor grid* (no. 3). The remaining three terminals are called the *controlled grid* (no. 1), the *plate,* and the *cathode,* respectively. The characteristic curves corresponding to a typical screen grid voltage V_{G_2} and suppressor grid voltage V_{G_3} are shown in Fig. 3-10*b* and *c*.

3-5-2 MULTITERMINAL ELEMENTS WITH PRESCRIBED CONSTRAINTS

There is a special, but very common, class of *n*-terminal black box which need not be operated as a three-terminal controlled resistor. This class of black box must possess certain prescribed constraints which allow a simple, yet complete, characterization to be possible. The prescribed constraints may take the following two common forms: (1) internal current and voltage constraints and (2) external port constraints.

Internal current and voltage constraints This type of black box is usually characterized by the property that one (or more) terminal currents or voltages are constrained to be always (or to be exact, approximately) zero. In addition to this, the black box is usually constrained by some internal elements which behave like controlled sources, such that certain terminal voltages are independent of their terminal currents. For example, a voltage-controlled voltage source can be considered as a four-terminal black box in which the current entering the controlling terminal is constrained to be zero, and the output voltage is constrained to be a function of the controlling voltage only.



(*a*) (*b*) (*c*)

Fig. 3-10. A pentode, when operating as a three-terminal resistor, can be characterized by two sets of characteristic curves by fixing the values of the screen grid voltage V_{G_2} and the suppressor grid voltage V_{G_3}.

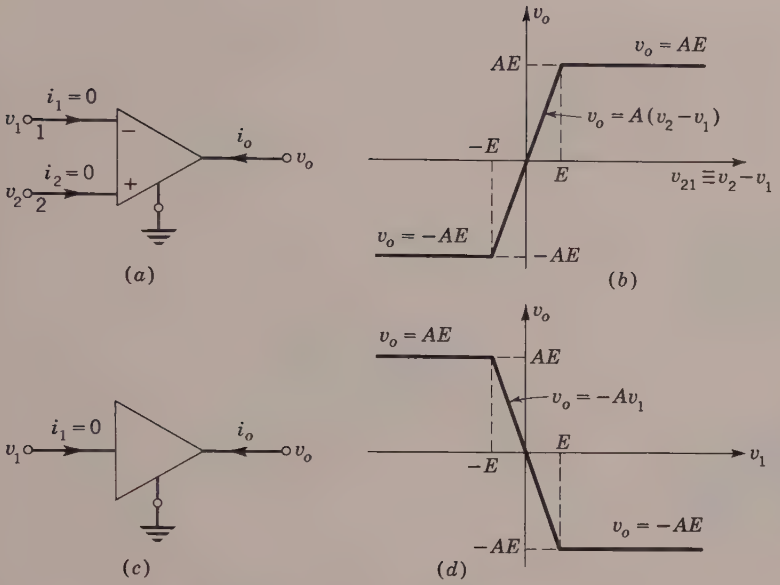


Fig. 3-11. An operational amplifier is a multiterminal resistor with a simple relationship between the output voltage v_o and the voltage v_{21} between terminals 1 and 2.

A widely used four-terminal resistor with internal current and voltage constraints is the *operational amplifier* whose symbol is shown in Fig. 3-11*a*.[1] This resistor is characterized by the following three relationships:

$$i_1 = 0$$

$$i_2 = 0$$

and

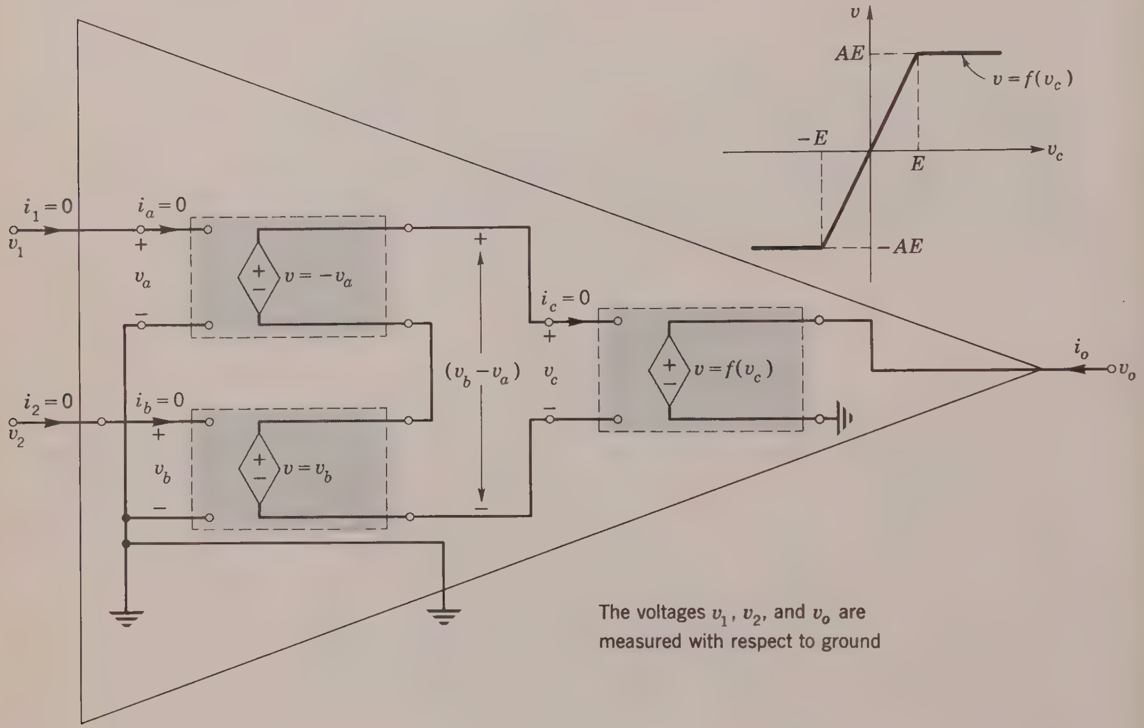
$$v_o = \begin{cases} A(v_2 - v_1) & \text{if } |v_2 - v_1| < E \\ AE & \text{if } (v_2 - v_1) \geq E \\ -AE & \text{if } (v_2 - v_1) \leq -E \end{cases} \tag{3-43}$$

It is important to observe that the third relationship is independent of the terminal current i_o. Moreover, since it depends only on $v_{21} \equiv (v_2 - v_1)$, the voltage difference between terminals 1 and 2, it is more convenient to plot this relationship in terms of $(v_2 - v_1)$, as shown in Fig. 3-11*b*. This is clearly simpler than the usual procedure of plotting v_o versus v_1, with v_2 as a parameter, or vice versa.

In most commercially available operational amplifiers, the value of A is at least 20,000, and the value of E is usually less than 0.5 mv. This means that the maximum magnitude of the out-

---

[1] When there is no ambiguity, it is a common practice to delete the ground terminal from this symbol for simplicity. In any case, the terminal voltages v_1, v_2, and v_o are always measured with respect to the ground terminal.



The voltages v_1, v_2, and v_o are measured with respect to ground

Fig. 3-12. An operational amplifier can be modeled by three controlled sources.

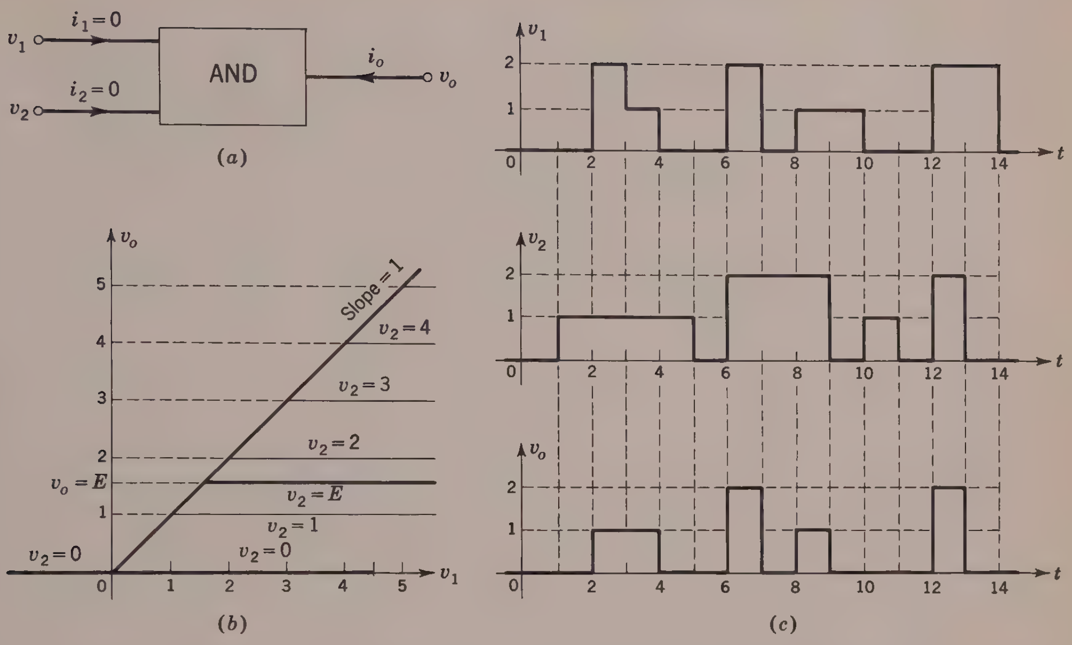
put voltage of most transistorized operational amplifiers is less than 10 volts.[1] In the less expensive type of operational amplifiers, there is only one input terminal, as shown in Fig. 3-11*c*. In this case, $v_2 = 0$, and the output voltage relationship reduces to that shown in Fig. 3-11*d*.

An operational amplifier can be modeled by three controlled sources, as shown in Fig. 3-12. This is easily seen to be true because $i_1 = i_a = 0$, $i_2 = i_b = 0$, and $v_c = v_b - v_a = v_2 - v_1$. In practice, the operational amplifier is constructed by an appropriate circuit (containing transistors, FET, etc.) which need not concern us now. As a matter of fact, with the advent of integrated-circuit technology, the internal elements which made up the operational amplifier have now been miniaturized so that as far as the user is concerned, he could hardly tell the difference between this element and a transistor. In other words, from a black-box point of view, the operational amplifier is just another multiterminal resistor.

The phenomenal progress in integrated-circuit technology has opened up an entirely new era in nonlinear circuits. There are now many commercially available multiterminal black boxes which be-

---

[1] The parameters A and E are usually called the *voltage gain* and the *saturation* voltage, respectively. The amplifier is said to be *saturated* whenever $|v_2 - v_1| > E$ because the output voltage ceases to vary.



(a)

(b)

(c)

Fig. 3-13. An AND gate is characterized by the property that the output voltage at any time t is equal to the smaller of the input voltages at the same time.

long to the above category. Most of these elements are called *logic elements* because they are the building blocks of systems, such as digital computers, which perform logic operations. We shall briefly mention two common logic elements, the *AND gate* and the *OR gate.*

The symbol of a four-terminal AND gate is shown in Fig. 3-13*a*.[1] This element is characterized by the relationships

$$i_1 = 0$$

$$i_2 = 0$$

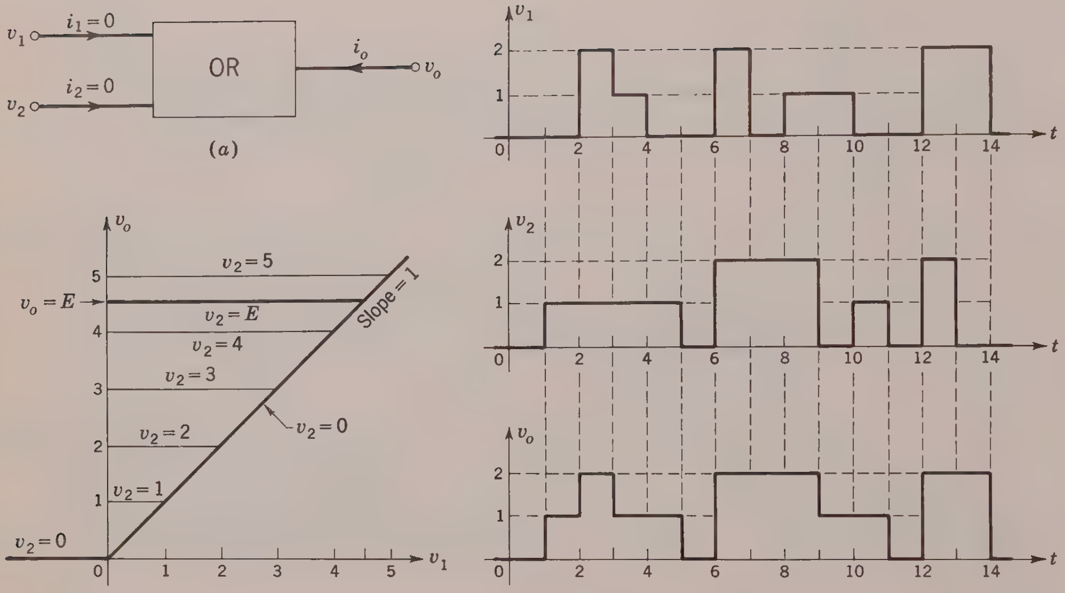
and

$$v_o = \begin{cases} v_1 & \text{whenever } 0 < v_1 \leq v_2 \\ v_2 & \text{whenever } 0 < v_2 \leq v_1 \\ 0 & \text{otherwise} \end{cases} \tag{3-44}$$

The third relationship can be represented by the curves shown in Fig. 3-13*b*. Observe that at any time t, the output voltage is equal to the smaller of the two input voltages, provided v_1 and v_2 are positive. Hence, if v_1 and v_2 consist of stepwise positive waveforms as shown in Fig. 3-13*c*, then the output voltage $v_o(t)$ will be

---

[1] For simplicity, the ground terminal is not shown. In any case, the voltages are measured with respect to the ground terminal.



(b) (c)

Fig. 3-14. An OR gate is characterized by the property that the output voltage at any time t is equal to the larger of the input voltages at the same time.

positive only if both v_1 *and* v_2 are positive. The conjunction AND is what gives the black box its name. An AND gate is, of course, not restricted to only four terminals. The synthesis of an n-terminal AND gate will be given in Chap. 10.

The symbol of a four-terminal OR gate is shown in Fig. 3-14*a*. This element is characterized by the relationships

$$i_1 = 0$$

$$i_2 = 0$$

and

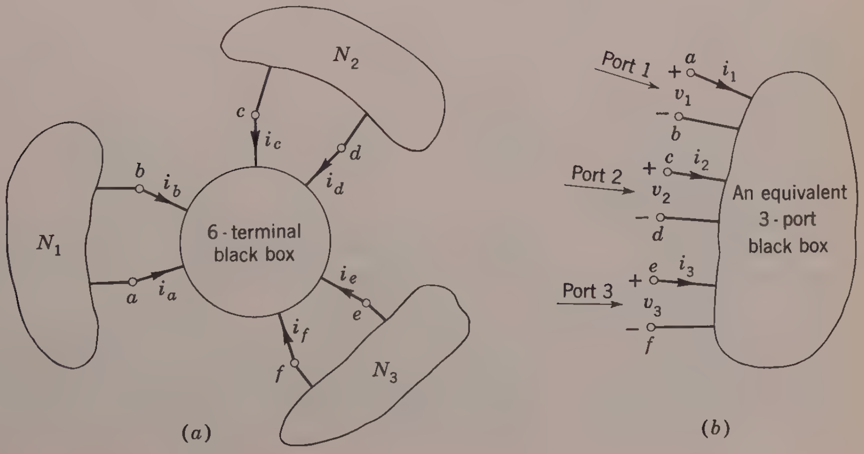
$$v_o = \begin{cases} v_1 & \text{whenever } v_1 \geq v_2 > 0 \\ v_2 & \text{whenever } v_2 \geq v_1 > 0 \\ 0 & \text{otherwise} \end{cases} \tag{3-45}$$

The third relationship can be represented by the curves shown in Fig. 3-14*b*. Observe that at any time t, the output voltage is equal to the larger of the two input voltages, provided v_1 and v_2 are positive. Hence, if v_1 and v_2 consist of stepwise positive waveforms as shown in Fig. 3-14*c*, then the output voltage $v_o(t)$ will be positive whenever v_1 *or* v_2 is positive. The conjunction OR is what gives the black box its name. An OR gate is, of course, not restricted to only four terminals. The synthesis of an n-terminal OR gate will be given in Chap. 10.

We shall conclude this section with the interesting observation that when the characteristics of the AND gate (Fig. 3-13*b*) and the OR gate (Fig. 3-14*b*) are plotted *on the same plane,* the entire first quadrant becomes filled up with a system of uniformly spaced horizontal lines, with the 45° line dividing the two sets of curves. Intuitively, this suggests their functions to be complementary to each other.

External port constraints This type of black box is usually constrained by external two-terminal networks so that the current entering one terminal is equal to the current leaving another terminal. For example, consider the six-terminal black box and the three two-terminal networks N_1, N_2, and N_3, as shown in Fig. 3-15*a*. It will be shown in the next chapter that so long as N_1, N_2, and N_3 are isolated from each other (i.e., no terminals inside N_1, N_2, and N_3 are connected to the external networks), then $i_a = -i_b$, $i_c = -i_d$, and $i_e = -i_f$. This is equivalent to saying that the currents entering terminals *a*, *c*, and *e* are equal to the currents leaving terminals *b*, *d*, and *f*, respectively. Any pair of terminals which satisfies this property is called a *port.* Hence, the six-terminal black box of Fig. 3-15*a* is also a three-port black box, and may be redrawn into the more explicit form shown in Fig. 3-15*b*. The significance of this interpretation is that instead of considering six current and six voltage variables, we need only three port-current

Fig. 3-15. A six-terminal black box can be constrained by three external two-terminal circuits to obtain a three-port black box.



(*a*) (*b*)

variables and three port-voltage variables. In general, an n-terminal black box can be operated as an $(n/2)$-port black box provided the current entering one terminal of each port is equal to the current leaving the other terminal of the same port. One simple method to satisfy this port constraint requirement is to connect an isolated two-terminal network across each port of the black box.

There are certain black boxes for which the port constraint requirement is automatically satisfied. An example of a two-port black box with this property is the *two-port ideal transformer* shown in Fig. 3-16*a*. This element is a simple model of an iron-core transformer consisting of two *separate* windings. The physical separation in this case guarantees that any current entering one terminal of each winding must leave the other terminal. A two-port ideal transformer is characterized by the relationship

$$\begin{aligned} v_1 &= n v_2 \\ i_1 &= -\frac{1}{n} i_2 \end{aligned} \tag{3-46}$$

where $n \equiv n_1/n_2$ is a constant called the *turns ratio.*

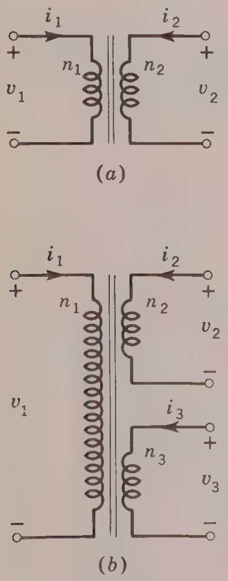


Fig. 3-16. Two common n-port black boxes which satisfy the port constraint requirement automatically.

Another n-port black box which satisfies the port constraint requirement automatically is the *three-port ideal transformer* shown in Fig. 3-16*b*. This element is a simple model of a multiwinding iron-core transformer. Again, the physical separation of the windings guarantees that the current entering one terminal of each port leaves the other terminal of the same port. This element is characterized by the relationship

$$\begin{aligned} i_1 &= -\frac{1}{n_1}(n_2 i_2 + n_3 i_3) \\ v_2 &= \frac{n_2}{n_1} v_1 \\ v_3 &= \frac{n_3}{n_1} v_1 \end{aligned} \tag{3-47}$$

Exercise 1: (*a*) Plot the output voltage characteristic curves v_o versus v_1 of an operational amplifier with v_2 as the parameter. (*b*) Repeat (*a*) for the v_o-vs.-v_2 curves with v_1 as the parameter.

Exercise 2: (*a*) Plot the characteristics of an AND gate in the v_o-vs.-v_2 plane, with v_1 as the parameter. (*b*) Repeat (*a*) for an OR gate on the same plane.

Exercise 3: A unity turns ratio two-port ideal transformer is usually called an *isolation transformer.* (*a*) Show how an isolation transformer can be used

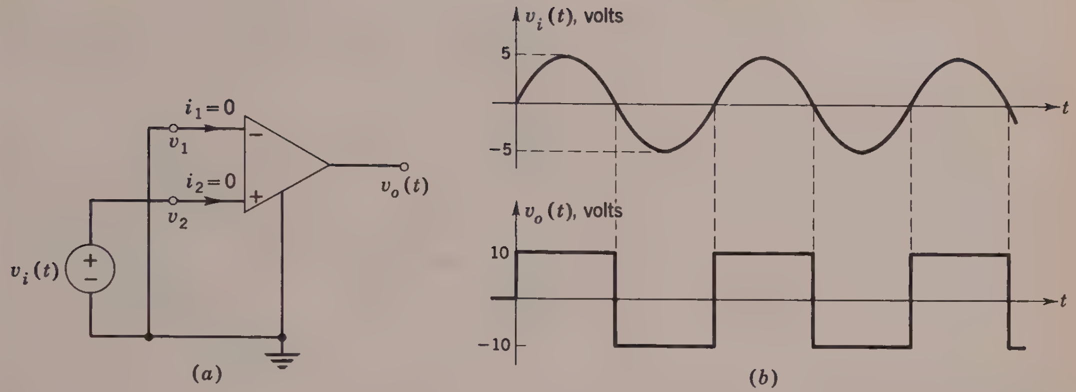


Fig. 3-17. A simple sine-wave-to-square-wave converter.

to guarantee any pair of terminals to be a "port" without altering the characteristics of any black box connected to it. (*b*) Show that any n-terminal black box (assume n is even) can be transformed into an $(n/2)$-port black box by using $n/2$ isolation transformers.

3-6 SOME PRACTICAL APPLICATIONS OF OPERATIONAL AMPLIFIERS

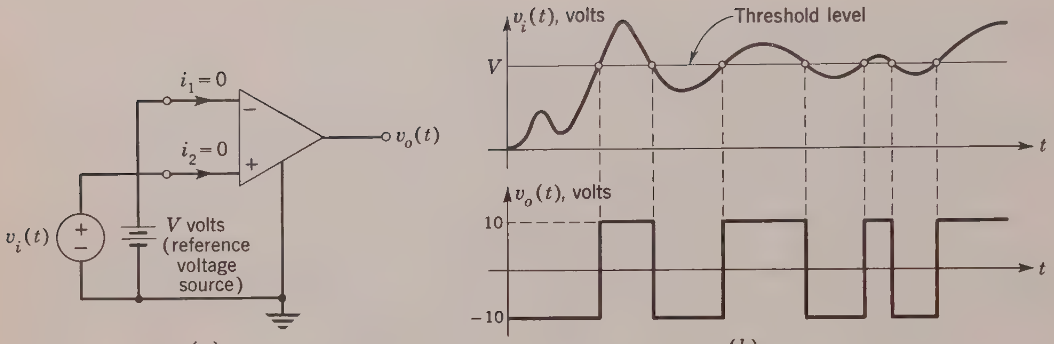
In terms of practical applications, the operational amplifier is one of the most versatile black boxes ever conceived.[1] An examination of the relationship shown in Fig. 3-11*b* shows that an operational amplifier is a nonlinear element if the magnitude of the voltage between terminals 1 and 2 exceeds the saturation voltage E. Otherwise, the operational amplifier will be operating in the linear region with the relationship $v_o = A(v_2 - v_1)$. Accordingly, there are two basic modes of applications: (1) nonlinear mode and (2) linear mode.

3-6-1 OPERATIONAL AMPLIFIER USED AS A NONLINEAR ELEMENT

For this type of application, the input voltages are applied directly across terminals 1 and 2 such that $|v_2 - v_1| > E$. Since the value of E for most practical operational amplifiers is less than 0.0005 volt, the operational amplifier is most certainly operating in the nonlinear *saturation* region for all input voltages greater than several millivolts. In this case, the value of E can be assumed to be zero for most practical purposes.

The simplest application of the operational amplifier operating in the nonlinear mode is that of converting a sine wave into an "almost" square wave. The simple circuit shown in Fig. 3-17*a* shows that $(v_2 - v_1) = v_2 - 0 = v_i(t)$. Hence, if $v_i(t) = V \sin \omega t$

[1] The operational amplifier was not used as extensively in the past as it is today, in spite of its versatility, because until the advent of integrated-circuit technology, it was a relatively bulky and expensive component. Both of these serious drawbacks have now been overcome, and the operational amplifier is fast becoming the workhorse of a large class of important linear and analog circuits, some of which will be discussed in this section.



(a) (b)

Fig. 3-18. A simple voltage comparator for producing a positive voltage whenever the input signal $v_i(t)$ exceeds a threshold level of V volts.

volts, then $v_o(t) = AE$ volts whenever $v_i(t) > E$, and $v_o(t) = -AE$ volts whenever $v_i(t) < E$. In many practical operational amplifiers, the typical value of (AE) is 10 volts, and the output voltage corresponding to a 5-volt sine wave is therefore as shown in Fig. 3-17*b*. Since laboratory square-wave generators are usually much more expensive than sine-wave generators, the above simple circuit will convert any sine-wave generator into a "poor man's square-wave generator." If the frequency of the square wave is unimportant, one can simply apply the line voltage and obtain a 60-Hz square wave.

Another simple application of the operational amplifier operating in the nonlinear mode is that of a *voltage comparator,* or *threshold detector*. In many automatic control systems it is necessary to compare the instantaneous voltage of an unknown input signal $v_i(t)$ with a fixed reference voltage $v_r(t) = V$ volts. It is desired to design a circuit that will "react" in some prescribed manner as soon as $v_i(t)$ attains the reference voltage V. A simple circuit for achieving this function is shown in Fig. 3-18*a*. Since $(v_2 - v_1) = v_i(t) - V$, the output voltage $v_o(t)$ will be positive (assuming $E \approx 0$) whenever $v_i(t) > V$ volts and negative whenever $v_i(t) < V$ volts. A typical input signal $v_i(t)$ and the corresponding output signal are shown in Fig. 3-18*b*. Since the reference voltage V can be interpreted as a "threshold," the circuit is also called a *threshold detector*. When $v_i(t)$ is obtained from the output of a transducer, a threshold detector is often used to provide a warning signal whenever the value of some physical controlling variable reaches a dangerous value. For example, in a nuclear reactor, a pressure transducer connected with a threshold detector can be used to prevent a catastrophic explosion, which did occur in the past, by providing the operator with a warning signal as soon as the internal pressure builds up above some preset level. In practice, the operator is usually a mechanism which will turn off the reactor automatically whenever such a dangerous pressure threshold is exceeded.

3-6-2 OPERATIONAL AMPLIFIER USED AS A LINEAR ELEMENT

In order to operate in its linear region, the magnitude of the voltage between terminals 1 and 2 of the operational amplifier must be less than the saturation voltage E, which is normally no larger than 0.0005 volt. Since the order of magnitude of input voltages used in applications involving operational amplifiers is usually much higher than 0.0005 volt, it is clear that in order to operate in the linear mode, the input voltages should not be applied directly to terminals 1 and 2. Instead, the operational amplifier is usually connected with a black box N_f and a black box N_c as shown in Fig. 3-19. The black box N_f feeds the output voltage v_o back into the input terminals, and is usually called the *feedback network.* The black box N_c "couples" the input signals into terminals 1 and 2 of the operational amplifier, and is usually called the *coupling network.* The effect introduced by these networks is that the voltage v_{12} between terminals 1 and 2 is no longer an independent voltage, which would be the case if the input voltages were applied directly across these terminals, but depends on the output voltage v_o and the networks N_f and N_c. In fact, since $v_o = A(v_2 - v_1) = -Av_{12}$ in the linear region,

$$v_{12} = \frac{-v_o}{A} \approx 0 \tag{3-48}$$

because for most practical operational amplifiers the value of A is at least 20,000, and the magnitude of v_o is less than 10 volts. Equation (3-48) is an extremely useful relationship to remember

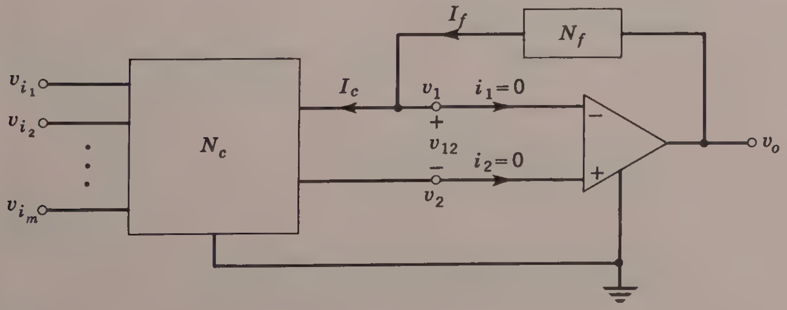


Fig. 3-19. In order to operate in the linear mode, the operational amplifier is usually connected to a feedback network N_f and a coupling network N_c.

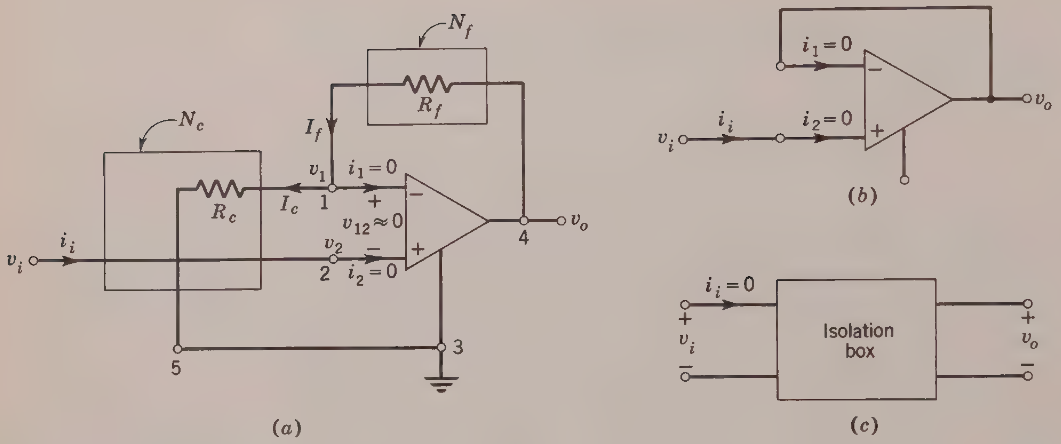


Fig. 3-20. An operational amplifier can be connected to function either (*a*) as a voltage-controlled voltage source with a positive controlling coefficient or (*b*) as an isolation box.

because it greatly simplifies the analysis of circuits containing operational amplifiers. Since a zero voltage is analogous to a short circuit, Eq. (3-48) is usually expressed as follows:[1]

PRINCIPLE OF VIRTUAL SHORT CIRCUIT

If the operational amplifier is connected to a feedback network N_f and a coupling network N_c such that the input voltages are not applied directly across terminals 1 and 2, then the voltage v_{12} can be set equal to zero for purposes of analysis.

The above principle will now be used to analyze the following practical circuits.

Voltage-controlled voltage source with positive controlling coefficients Consider the circuit shown in Fig. 3-20*a*. Applying Kirchhoff current and voltage laws, we obtain

KCL at terminal 1

$$I_c = I_f \tag{3-49}$$

KVL at loop around terminals 4-1-5-3-4

$$R_f I_f + R_c I_c - v_o = 0 \tag{3-50}$$

KVL at loop around terminals 1-5-3-2-1

$$R_c I_c - v_i - v_{12} = 0 \tag{3-51}$$

---

[1] The adjective *virtual* is used to caution against our making the common mistake of connecting terminals 1 and 2 together with an actual short circuit, in order to simplify analysis. This is illegal because a current will flow down this short circuit, thereby altering the original relationship $I_c = I_f$. On top of this, v_{12} is not really zero; it is only approximately zero when compared with the other terminal voltages. Observe also that $v_{12} = 0$ would have rendered the circuit useless because $v_o = -Av_{12}$ would then be zero.

Substituting Eq. (3-49) for I_f in Eq. (3-50) and $v_{12} = 0$ in Eq. (3-51), we obtain

$$v_o = (R_f + R_c)I_c \tag{3-52}$$

$$v_i = R_c I_c \tag{3-53}$$

Dividing Eq. (3-52) by Eq. (3-53), we obtain

$$v_o = \left(\frac{R_f + R_c}{R_c}\right) v_i \tag{3-54}$$

Since the input current $i_i = i_2 = 0$, the resulting network is simply a voltage-controlled voltage source with a positive controlling coefficient k given by

$$k = \frac{R_f + R_c}{R_c} \tag{3-55}$$

Observe that in the special case where $R_f = 0$ (short circuit) and $R_c = \infty$ (open circuit), the circuit reduces to that shown in Fig. 3-20*b*, and Eq. (3-55) reduces to $k = 1$. This circuit is now described by

$$\begin{aligned} v_o &= v_i \\ i_i &= 0 \end{aligned} \tag{3-56}$$

and is called an *isolation box* because it serves no purpose other than to isolate the external circuit connected across the output terminals from (drawing any current) the circuit connected across the input terminals. We shall denote an isolation box by the symbol shown in Fig. 3-20*c*.[1]

Voltage-controlled voltage source with negative controlling coefficients Consider the circuit shown in Fig. 3-21*a*. Applying Kirchhoff current and voltage laws, we obtain

KCL at terminal 1

$$I_c = I_f \tag{3-57}$$

KVL at loop around terminals 4-1-2-5-3-4

$$R_f I_f + v_{12} - v_o = 0 \tag{3-58}$$

---

[1] In practice, an isolation box is sometimes called an *emitter follower* or a *cathode follower*, depending on whether the input terminal is connected with the emitter of a transistor or the cathode of a vacuum tube. This terminology comes from the interpretation that the output voltage "follows" the input voltage ($v_o \approx v_i$) in a practical isolation box.

KVL at loop around terminals 1-6-3-5-2-1

$$R_c I_c + v_i - v_{12} = 0 \tag{3-59}$$

Substituting Eq. (3-57) into Eq. (3-58) and assuming $v_{12} = 0$, we obtain, upon dividing Eq. (3-58) by Eq. (3-59),

$$v_o = -\frac{R_f}{R_c} v_i \tag{3-60}$$

Observe that the input current

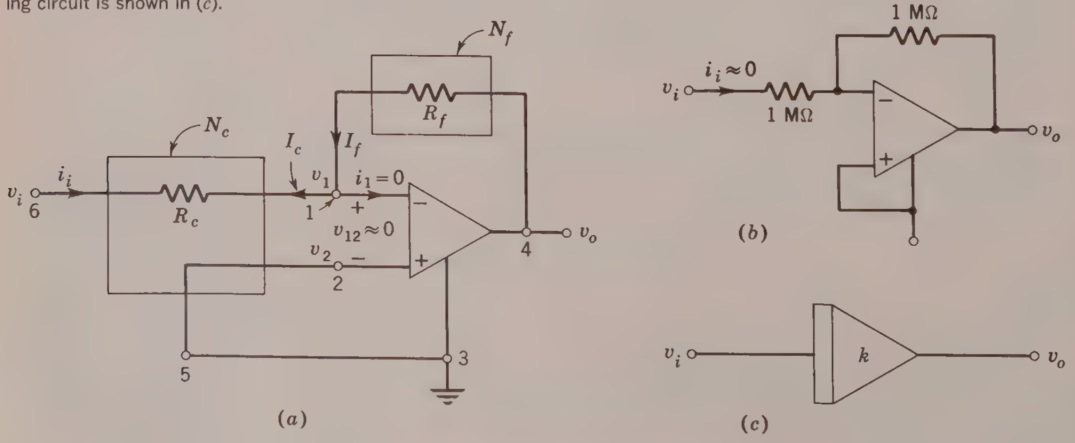
$$i_i = \frac{v_i}{R_c} \tag{3-61}$$

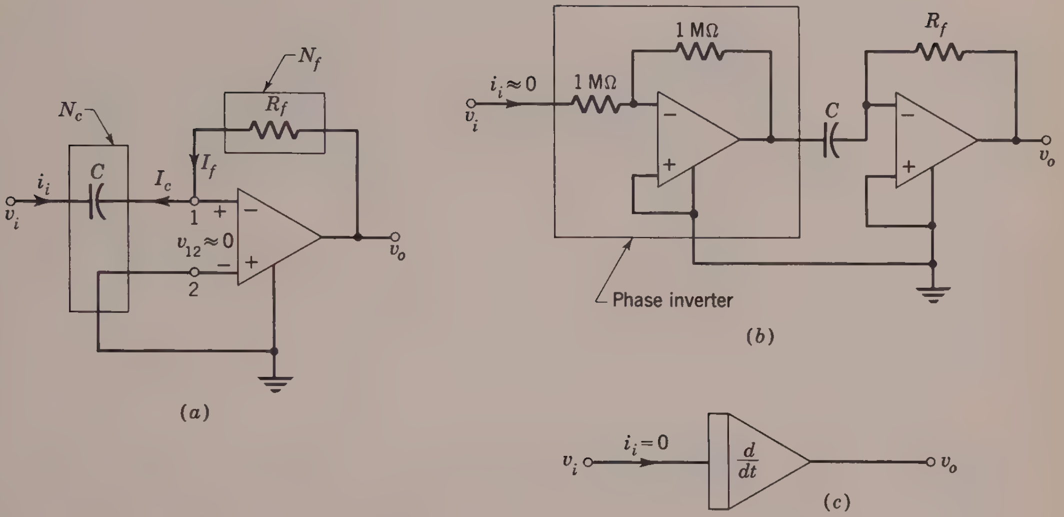
is not zero. However, if we choose the value of R_c to be very large (say 1 MΩ), then $i_i \approx 0$ for most practical purposes, and the circuit becomes a voltage-controlled voltage source with a negative controlling coefficient k given by

$$k = -\frac{R_f}{R_c} \tag{3-62}$$

In the special case where $R_f = R_c \geq 1$ MΩ, the circuit reduces to that shown in Fig. 3-21*b*. This circuit is usually called a *phase inverter* because the output voltage $v_o(t)$ corresponding to an input sinusoidal voltage is the same sinusoidal waveform inverted about the time axis; i.e., the phase is changed by 180°. Hence, between

Fig. 3-21. An operational amplifier can be connected to function either as (*a*) a voltage-controlled voltage source with a negative controlling coefficient or (*b*) as a phase inverter. The general symbol of a scaling circuit is shown in (*c*).





(*a*)

(*b*)

(*c*)

Fig. 3-22. (*a*) An operational amplifier can be connected to convert an input voltage waveform into an output voltage waveform proportional to its time derivative. (*b*) If a phase inverter is connected across the input terminals, the result is called a differentiator. The symbol of a differentiator is shown in (*c*).

the two circuits shown in Figs. 3-20*a* and 3-21*a*, we can multiply or divide an input voltage $v_i(t)$ by either a positive or a negative constant. We shall refer to these two general circuits as a *scaling circuit* and shall denote it by the symbol shown in Fig. 3-21*c*.[1] The constant k inside this symbol is the scaling constant.

Differentiator If we replace the resistor R_c in Fig. 3-21*a* with a linear capacitor with capacitance equal to C farads, as shown in Fig. 3-22*a*, then applying the same procedure as before, we obtain

$$v_o = R_f I_c \tag{3-63}$$

and

$$v_i(t) = -\frac{1}{C}\int_{-\infty}^{t} I_c(\tau)\,d\tau \tag{3-64}$$

If we differentiate both sides of Eq. (3-64) with respect to time, we obtain

$$C\frac{dv_i(t)}{dt} = -I_c(t) \tag{3-65}$$

Dividing Eq. (3-63) by Eq. (3-65), we obtain

$$v_o(t) = -R_f C\frac{dv_i(t)}{dt} \tag{3-66}$$

---

[1] In this symbol, and in several others in the following pages, the ground terminal is omitted for convenience. In all cases, the voltages are measured with respect to a common ground.

Equation (3-66) shows that the output voltage is proportional to the time derivative of the input voltage. Since the input current i_i is generally not zero, and since the constant of proportionality is negative, it is often desirable to connect the phase inverter circuit of Fig. 3-21*b* across the input terminal of this circuit, as shown in Fig. 3-22*b*. The resulting black box is called a *differentiator* because

$$v_o(t) = k\frac{dv_i(t)}{dt} \tag{3-67}$$

and

$$i_i(t) = 0 \tag{3-68}$$

The constant $k = R_f C$ can be set equal to any desired value, including unity, by an appropriate choice of values for R_f and C. The usual symbol of a differentiator is shown in Fig. 3-22*c*. In applications where the sign of k is immaterial, we can save one operational amplifier by choosing a very large resistance for R_f. In this case, the current I_f, and hence I_c, can be made arbitrarily small so as to satisfy the condition given by Eq. (3-68).

Integrator If we replace the resistor R_f in Fig. 3-21*a* with a linear capacitor with capacitance equal to C farads, as shown in Fig. 3-23*a*, then applying the same procedure as before, we obtain

$$v_o(t) = \frac{1}{C}\int_{-\infty}^{t} I_f(\tau)\,d\tau \tag{3-69}$$

$$v_i = -R_c I_f \tag{3-70}$$

Solving for I_f in Eq. (3-70) and substituting the result in Eq. (3-69), we obtain

$$v_o(t) = -\frac{1}{R_c C}\int_{-\infty}^{t} v_i(\tau)\,d\tau \tag{3-71}$$

Equation (3-71) shows that the output voltage is proportional to the time integral of the input voltage. Since the input current i_i is generally not zero, and since the constant of proportionality is negative, it is often desirable to connect a phase inverter across the input terminals of the circuit as shown in Fig. 3-23*b*. The resulting black box is called an *integrator* because

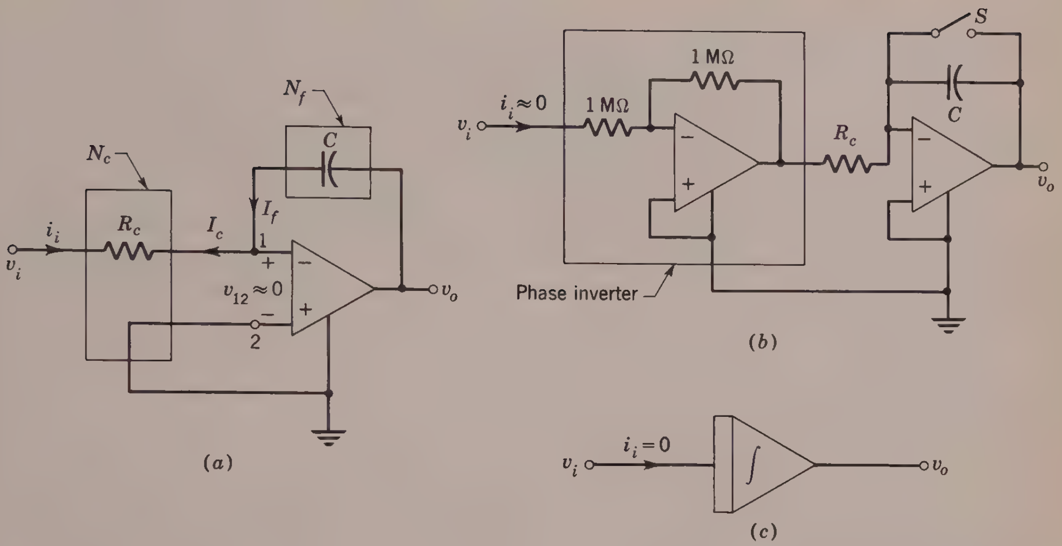


Fig. 3-23. (*a*) An operational amplifier can be connected to convert an input voltage waveform into an output voltage waveform proportional to its time integral. (*b*) If a phase inverter is connected across the input terminals, the result is called an integrator. The symbol of an integrator is shown in (*c*).

$$v_o(t) = k \int_{-\infty}^{t} v_i(\tau)\, d\tau \tag{3-72}$$

and

$$i_i = 0 \tag{3-73}$$

The constant $k = R_cC$ can be set equal to any desired value, including unity, by an appropriate choice of values for R_c and C. Observe that the lower limit of integration in Eq. (3-72) must be set equal to $-\infty$ because $v_i(t)$ may not be zero for $t < 0$. In practice, it is common to connect a switch S across the capacitor as shown in Fig. 3-23*b*. When the integration is desired for all times $t \geq t_0$, the switch, which is normally closed, is opened at $t = t_0$, and Eq. (3-72) reduces to

$$v_o(t) = k \int_{t_0}^{t} v_i(\tau)\, d\tau \tag{3-74}$$

The usual symbol of an integrator is shown in Fig. 3-23*c*. Observe that in applications where the sign of k is unimportant, we can save one operational amplifier by choosing a high resistance value for R_C (see Appendix C for a practical integrator circuit).

Summing circuit Consider next the circuit shown in Fig. 3-24*a*, where the coupling network N_c consists of a number of resistors

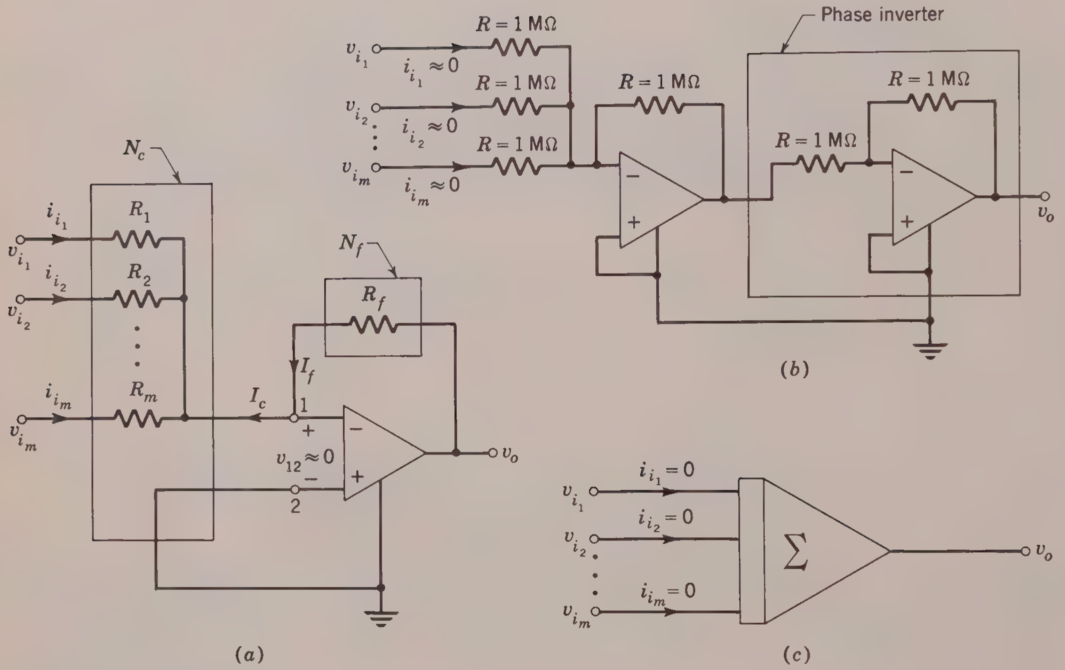


Fig. 3-24. (*a*) An operational amplifier can be connected to give an output voltage which is a linear combination of the input voltages; (*b*) if a phase inverter is connected across the output terminals, the result is called a *summing circuit.* The symbol of a summing circuit is shown in (*c*).

$R_1, R_2, \ldots, R_m$. Since $v_{12} = 0$, the voltage across each resistor R_j is equal to v_{i_j}. Therefore, applying the Kirchhoff current law at the junction point of these resistors, we obtain

$$I_c = -\left[\frac{v_{i_1}}{R_1} + \frac{v_{i_2}}{R_2} + \cdots + \frac{v_{i_m}}{R_m}\right] \tag{3-75}$$

The output voltage is given as before by

$$v_o = R_f I_f = R_f I_c \tag{3-76}$$

Substituting Eq. (3-75) into Eq. (3-76), we obtain

$$v_o = -\left[\frac{R_f}{R_1} v_{i_1} + \frac{R_f}{R_2} v_{i_2} + \cdots + \frac{R_f}{R_m} v_{i_m}\right] \tag{3-77}$$

Hence, apart from the negative sign, the output voltage is a "weighted" sum of the input voltages, with R_f/R_j as the weighting factor for each input voltage v_{i_j}. There are practical applications where this weighting feature is desirable. Observe that since

$$i_{i_j} = \frac{v_{i_j}}{R_j} \tag{3-78}$$

it is desirable to choose a high resistance value for R_j in order that $i_{i_j} \approx 0$. In practice, it is usually desirable to get rid of the negative sign in Eq. (3-77) as well as to have a pure summing operation. This can be achieved by choosing the same resistance values for all resistors (at least 1 MΩ to ensure $i_{i_j} \approx 0$) and by connecting a phase inverter across the output terminals as shown in Fig. 3-24*b*. In this case, it is easy to see that

$$v_o = v_1 + v_2 + \cdots + v_m \tag{3-79}$$

and

$$i_{i_j} \approx 0 \qquad j = 1, 2, \ldots, m \tag{3-80}$$

This circuit is called a *summing circuit* and is usually denoted by the symbol shown in Fig. 3-24*c*.

The various circuits shown in Figs. (3-20) to (3-24) are generally referred to as *analog circuits* because they can be used to *simulate* physical systems described by analogous mathematical equations. It is easy to see that the basic mathematical operations which can be simulated are (1) multiplication and division of a variable by a positive or a negative constant, (2) differentiation and integration of a variable with respect to time, and (3) addition and subtraction of two or more variables which are functions of time.

In order to illustrate the concept of simulation, consider the familiar spring-mass system and its driving force $f(t)$ as shown in Fig. 3-25. If we assume a spring constant equal to k and a friction force proportional to the velocity (with coefficient of friction equal to μ), then from freshman physics we know that the equation of motion of this system is given by

$$m\frac{d^2x}{dt^2} + \mu\frac{dx}{dt} + kx = f(t) \tag{3-81}$$

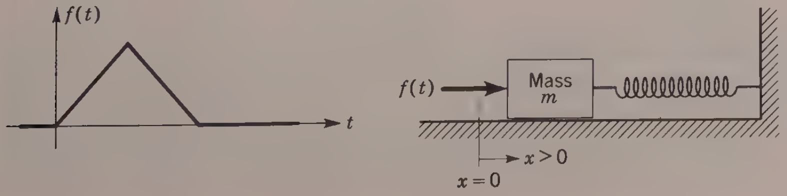


Fig. 3-25. A typical spring-mass system under an external driving force $f(t)$.

Our object is to investigate the resulting motion of this system corresponding to different values of m, μ, and k. Equation (3-81) is an example of a *linear differential equation,* and may be solved by various mathematical techniques which need not concern us now. Suffice it to say that for the given triangular forcing waveform $f(t)$, the analytical methods for solving this equation are rather tedious. It would be impractical, therefore, to repeat the methods for many different values of m, μ, and k. However, we can easily simulate this system if we represent x by the voltage variable v and rewrite Eq. (3-81) into the following more convenient form:

$$\frac{d^2v}{dt^2} = -\frac{\mu}{m}\frac{dv}{dt} - \frac{k}{m}v + \frac{1}{m}f(t) \tag{3-82}$$

To simulate this system, we start by applying a voltage equal to $d^2v(t)/dt^2$ at the input (point a) of an integrator, thereby obtaining the voltage $dv(t)/dt$ at point b. If we connect point b to the input of another integrator, the output voltage at point c will be $v(t)$. Let us next connect a scaling circuit with a negative scaling coefficient $-k_1 = -\mu/m$ and $-k_2 = -k/m$ at points b and c, respectively. The corresponding output voltages at points e and d will be $-(\mu/m)[dv(t)/dt]$ and $-(k/m)v(t)$, respectively. Applying these voltages and a voltage source with terminal voltage equal to $(1/m)f(t)$ to the input terminals of a summing circuit, we obtain at point f the output voltage

$$v_f = -\frac{\mu}{m}\frac{dv(t)}{dt} - \frac{k}{m}v(t) + \frac{1}{m}f(t) \tag{3-83}$$

Comparing Eq. (3-82) with Eq. (3-83), we observe that if we close the switch S, thereby making the voltage at point a equal to the voltage at point f, the resulting system would have exactly the same equations of motion as Eq. (3-82). Since this is an actual circuit in operation, we can observe the voltage $v(t)$ with the help of an oscilloscope. The resulting waveform for $v(t)$ is obviously the desired solution. It is now a simple matter to adjust the value of the coefficients k_1 and k_2 systematically, each time observing the resulting waveform. A photograph of each waveform can be taken for comparison later.

The above example is a simple demonstration of how to analyze nonelectrical systems by simulation. The circuit arrangement shown in Fig. 3-26 is called an *analog computer.* Although

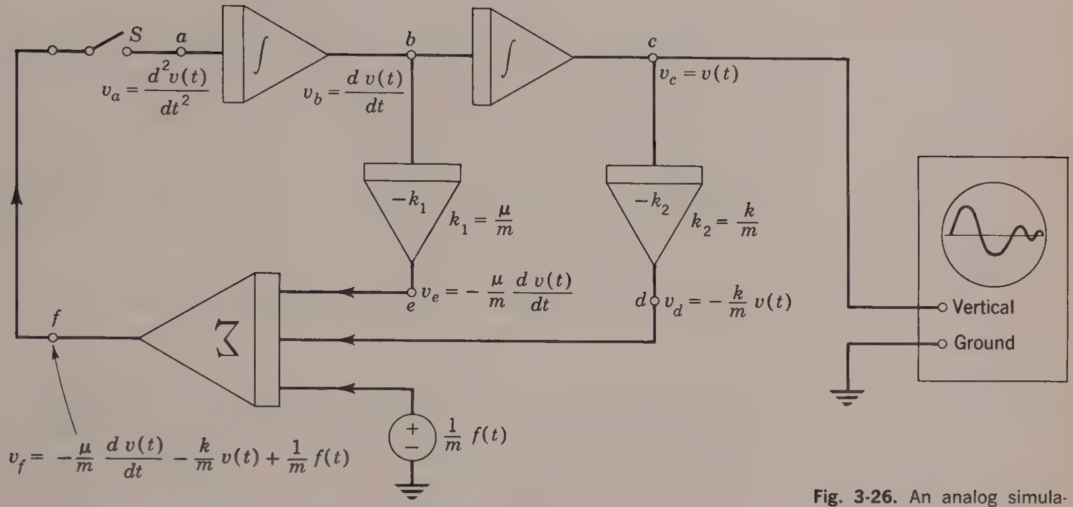

S
a
$v_a = \frac{d^2 v(t)}{dt^2}$
b
$v_b = \frac{d\,v(t)}{dt}$
c
$v_c = v(t)$
$-k_1$
$k_1 = \frac{\mu}{m}$
$-k_2$
$k_2 = \frac{k}{m}$
e
$v_e = -\frac{\mu}{m}\frac{d\,v(t)}{dt}$
d
$v_d = -\frac{k}{m}v(t)$
f
Σ
$v_f = -\frac{\mu}{m}\frac{d\,v(t)}{dt} - \frac{k}{m}v(t) + \frac{1}{m}f(t)$
$\frac{1}{m}f(t)$
Vertical
Ground

Fig. 3-26. An analog simulation of the spring-mass system of Fig. 3-25.

the commercial analog computer is a much more sophisticated machine, its principle of operation is basically the same. Essentially, a commercial analog computer is simply a collection of many identical analog circuits similar to those described earlier. Using these readily available units, the user can simulate many physical systems much more complex than the example above.

Finally, we want to point out that there are certain practical problems associated with circuits containing operational amplifiers which we did not mention previously. The most serious problems are the offset voltage, the drift, and the frequency limitation. In the ideal operational amplifier, the output voltage must be zero whenever the input voltage is zero. In practice, however, a small "offset voltage" is usually present, and some means must be taken to compensate for it. Another practical problem is that the characteristics of the operational amplifier tend to drift with temperature, time, and power supply voltages. The third problem is that most practical operational amplifiers do not work properly at very high frequencies. Again, this departure from the ideal performance is caused primarily by parasitic elements.

Exercise 1: Consider the square-wave converter shown in Fig. 3-17. To show that the steepness of the square wave depends on both the amplitude and frequency of the input waveform, plot one period of $v_o(t)$ corresponding to a 60-Hz sine wave with the following amplitudes: (*a*) 10 mv, (*b*) 5 mv, and (*c*) 1 mv. Assume that $E = 0.5$ mv and $A = 20{,}000$.

Exercise 2: The threshold detector circuit shown in Fig. 3-18*a* is used aboard a spacecraft to detect an abnormal rise in cabin temperature. If the temperature transducer output voltage during a certain test flight is given by $v_i(t) = 10 - 10e^{-t} \cos \pi t$ volts, plot the output voltage $v_o(t)$ corresponding to a preset threshold voltage $V = 10.5$ volts. Find the time intervals for which the cabin temperature exceeds the threshold temperature.

Exercise 3: Consider the scaling circuits shown in Figs. 3-20 and 3-21. Let the largest available resistance be equal to 10 MΩ. (*a*) Find the maximum possible magnitude of the scaling coefficient k in each case. Assume that $R_c = 1$ MΩ. (*b*) Find the minimum input current i_i flowing into a $k = 1{,}000$ and $k = -100$ scaling circuit when $v_i(t) = 10$ volts.

Exercise 4: Why is it desirable to have approximately zero input currents for the analog circuits described above? How can an isolation box be used to achieve this objective?

Exercise 5: An alternate method for realizing a differentiator or an integrator consists of replacing one of the two resistors in Fig. 3-21*a* by an inductor. Show how this can be done and derive the necessary equations to support your answer.

Exercise 6: (*a*) Find the function performed by the circuit shown in Fig. 3-20*a* when R_f is replaced by a linear capacitor. (*b*) Repeat (*a*) with R_c replaced by a linear capacitor.

Exercise 7: Draw an analog circuit for subtracting two voltage waveforms $v_1(t)$ and $v_2(t)$.

Exercise 8: Show that the two scaling circuits shown in Fig. 3-26 are superfluous if the summing circuit of Fig. 3-24*a* is used. Specify the corresponding weighting factors.

Exercise 9: (*a*) Show that it is possible to generate a periodic triangular waveform of any frequency (within the operating frequency limits of the operational amplifiers) by connecting the output of a sine-wave–to–square-wave converter across an integrator. (*b*) Find the relationship between the amplitude and frequency of the triangular output waveform and the sinusoidal input waveform.

Exercise 10: Show that with the use of only one operational amplifier, two identical resistors, and two identical capacitors, it is possible to design an integrator circuit whose output voltage is the time integral of the difference between two input voltage waveforms.

3-7 SCALORS, ROTATORS, AND REFLECTORS: A CLASS OF USEFUL TWO-PORT RESISTORS[1]

The functions performed by the operational amplifier circuits discussed in the preceding section can be summarized in one statement: They transform input voltage waveforms (functions of time) into some desired output voltage waveforms (functions of time). It is important to observe that the independent variable of the transformation is always the time *t*. Since the input currents

---

[1] A more detailed presentation of these devices is given in L. O. Chua, The Rotator–A New Network Component, *Proc. IEEE,* vol. 55, no. 9, pp. 1566–1577, September, 1967; and L. O. Chua, Synthesis of New Nonlinear Network Elements, *Proc. IEEE,* vol. 56, no. 8, pp. 1325–1340, August, 1968.

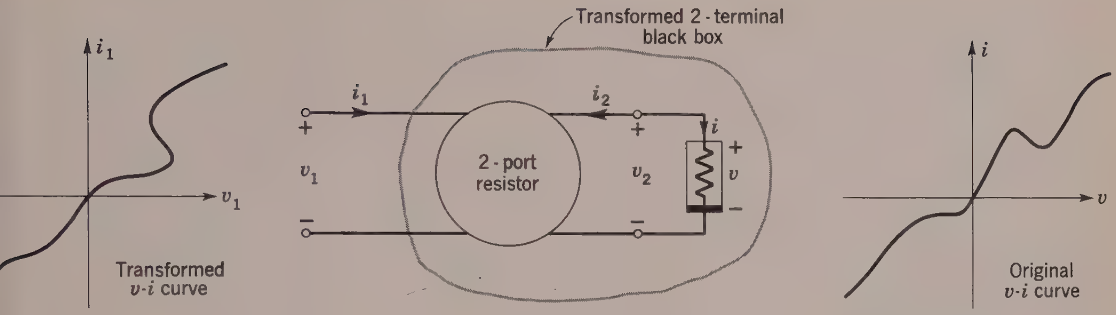


Fig. 3-27. The v-i curve of a given nonlinear resistor is transformed into a new v-i curve by connecting the resistor across one port of a two-port resistor.

are assumed to be zero, the analog circuits can be arbitrarily connected in any desired manner without affecting the original function performed by each circuit.

There is another important class of networks which also performs certain transformations, but the independent variable is not time. This class of networks takes the form of a two-port black box, and is called a *two-port resistor* because each black box is characterized by two relationships in terms of the port variables v_1, v_2, i_1, and i_2. If we connect a nonlinear resistor across port 2 of a two-port resistor as shown in Fig. 3-27, the resulting two-terminal black box can be interpreted as a *new* nonlinear resistor because it will have a v_1-i_1 curve different from the original v-i curve. In other words, the function performed by the two-port resistor is that of transforming a given v-i curve into a new v_1-i_1 curve. In this sense, we can generate many new nonlinear resistors from those that are presently available commercially. Of course, an arbitrary transformation is not likely to do us much good. What we need is to discover a few basic transformations from which all others can be obtained. Amazingly, only three types of transformations are necessary, namely, a scaling transformation, a rotation transformation, and a reflection transformation. In the scaling transformation, the abscissa or the ordinate of each point on the v-i curve is multiplied by a positive constant k, as shown in Fig. 3-28*a*. In the rotation transformation, the original v-i curve is rotated through an angle θ with respect to the origin, as shown in Fig. 3-28*b*. Finally, in the reflection transformation, the original v-i curve is reflected (the mirror image) with respect to some straight line through the origin, as shown in Fig. 3-28*c*. It is possible to find three types of two-port resistors for implementing the above transformations. Accordingly, they are called *scalors, rotators,* and *reflectors,* respectively.

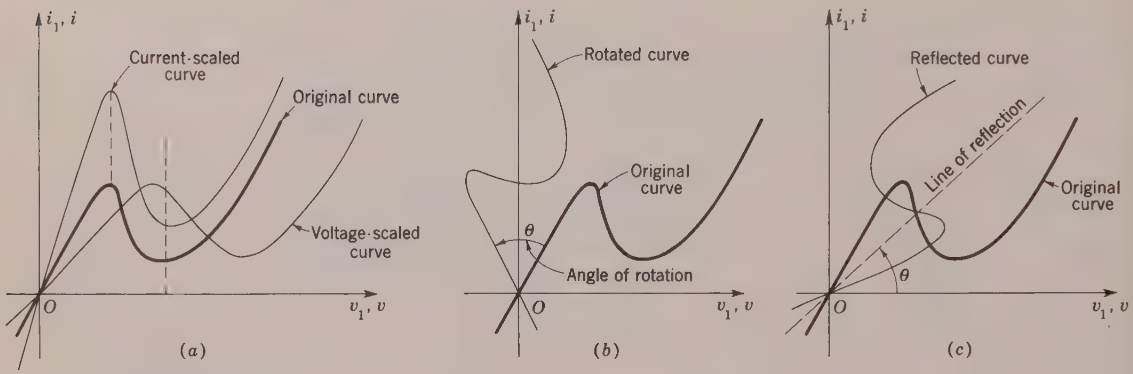


Fig. 3-28. A *v-i* curve may be (*a*) scaled with respect to the *v* axis or the *i* axis; (*b*) rotated with respect to the origin; or (*c*) reflected with respect to a line of reflection.

3-7-1 SCALORS

There are two types of scalors, *voltage scalors* and *current scalors.* As the name implies, a voltage scalor multiplies the voltage (abscissa) of each point on the *v-i* curve by a prescribed constant k_v, while maintaining the same value of current at the same point. This requirement can be characterized by

$$v_1 = k_v v_2 \tag{3-84}$$

and

$$i_1 = -i_2 \tag{3-85}$$

The negative sign in Eq. (3-85) is necessary in order to have $i_1 = i$.† We shall denote a voltage scalor by the symbol shown in Fig. 3-29*a*. Equations (3-84) and (3-85) can be easily synthesized by various controlled-source configurations. The reader can easily verify that the circuit shown in Fig. 3-29*b* and *c* constitutes two simple realizations.

By analogous reasoning, a current scalor multiplies the current (ordinate) of each point on the *v-i* curve by a prescribed constant k_i, while maintaining the same value of voltage at the same point. This requirement can be characterized by

$$v_1 = v_2 \tag{3-86}$$

and

$$i_1 = -k_i i_2 \tag{3-87}$$

---

† Throughout this section, the reference direction for i_1 and i_2 is assumed to enter the respective ports, as shown in Fig. 3-27.

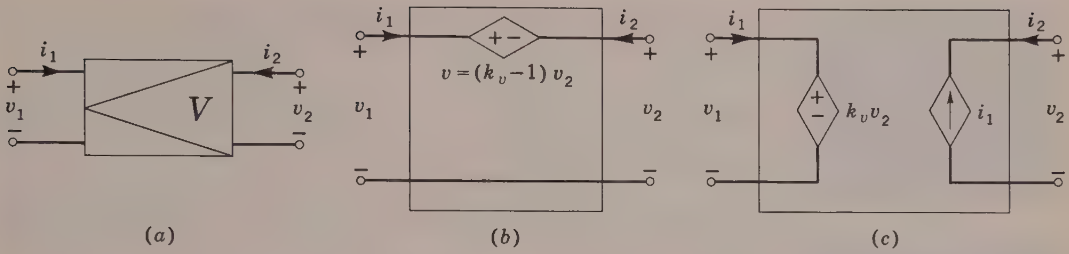


Fig. 3-29. A voltage scalor can be synthesized by either one or two controlled sources.

We shall denote a current scalor by the symbol shown in Fig. 3-30*a*. It is easy to see that the circuit shown in Fig. 3-30*b* and *c* represents two simple realizations of a current scalor.

The above basic controlled-source circuits can be synthesized in practice by various appropriate combinations of operational amplifiers or transistors. A practical circuit for each is given in Appendix C. For the sake of demonstration, the oscilloscope tracing of a typical v-i curve is shown in Fig. 3-31*a*. The corresponding v_1-i_1 curves obtained from a voltage scalor and a current scalor are shown in Fig. 3-31*b* and *c*, respectively.

It is important to realize that a scalor is completely different from the scaling circuits described in Sec. 3-6. The independent variable for a scalor is either a voltage or a current, whereas the independent variable for a scaling circuit is time. Since two relationships must be satisfied by a scalor, in comparison with only one in the scaling circuit, it is more difficult to realize a scalor in practice.

3-7-2 ROTATORS

From analytic geometry we know that the relationship required to rotate a point P with coordinates (v,i) into a point P' with

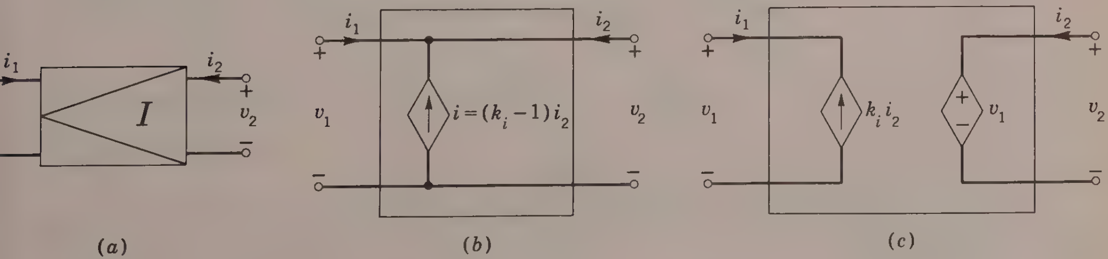


Fig. 3-30. A current scalor can be synthesized by either one or two controlled sources.

coordinates (v_1, i_1) by $\theta°$ (in the counterclockwise direction) is given by

$$v_1 = (\cos\theta)v - (\sin\theta)i$$

$$i_1 = (\sin\theta)v + (\cos\theta)i$$

Recasting these equations in terms of port current and voltage variables, we obtain

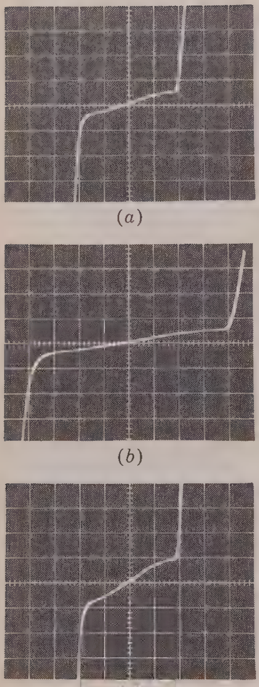
$$v_1 = (\cos\theta)v_2 + (\sin\theta)i_2$$

$$i_1 = (\sin\theta)v_2 - (\cos\theta)i_2$$

These equations assume that the v-i curve is drawn in the volt-ampere plane. In order to allow an arbitrary current scale, we multiply i_1 and i_2 in these equations by a scale factor R, thereby obtaining

$$v_1 = (\cos\theta)v_2 + (\sin\theta)(Ri_2) \tag{3-88}$$

$$i_1 = (\sin\theta)\frac{v_2}{R} - (\cos\theta)i_2 \tag{3-89}$$



(a)

(b)

(c)

Horizontal scale:
$v = 2.5$ volts per division
Vertical scale:
$i = 2$ ma per division

Fig. 3-31. (*a*) Actual oscilloscope picture of a typical v-i curve, the curve being transformed (*b*) by a voltage scalor into a voltage-scaled v-i curve and (*c*) by a current scalor into a current-scaled v-i curve.

Hence a rotator is completely characterized by Eqs. (3-88) and (3-89), where θ is the desired counterclockwise angle of rotation and R is a scale factor.[1] In the volt-milliampere plane, $R = 10^3$. We shall denote a rotator by the symbol shown in Fig. 3-32*a*. Equations (3-88) and (3-89) can be synthesized either by a π-network configuration of linear resistors as shown in Fig. 3-32*b* or by a T-network configuration as shown in Fig. 3-32*c*. In order to verify these realizations, we need only to derive v_1 and i_1 in terms of v_2 and i_2 for each network and show that they agree with Eqs. (3-88) and (3-89). For example, consider the π-network realization with conductances G_1, G_2, and G_3. By inspection, $i_a = G_1 v_1$, $i_b = G_3(v_1 - v_2)$, and $i_c = G_2 v_2$. Applying Kirchhoff current laws to each of the two junction points, we obtain

$$i_1 = G_1 v_1 + G_3(v_1 - v_2) = \left(\tan\frac{\theta}{2}\right)\frac{1}{R}v_1 + (-\csc\theta)\frac{1}{R}(v_1 - v_2) \tag{3-90}$$

$$i_2 = G_2 v_2 - G_3(v_1 - v_2) = \left(\tan\frac{\theta}{2}\right)\frac{1}{R}v_2 - (-\csc\theta)\frac{1}{R}(v_1 - v_2) \tag{3-91}$$

---

[1] A negative θ, therefore, corresponds to a clockwise angle of rotation.

With the help of the trigonometric identity

$$\tan\frac{\theta}{2} = (\csc\theta - \cot\theta)$$

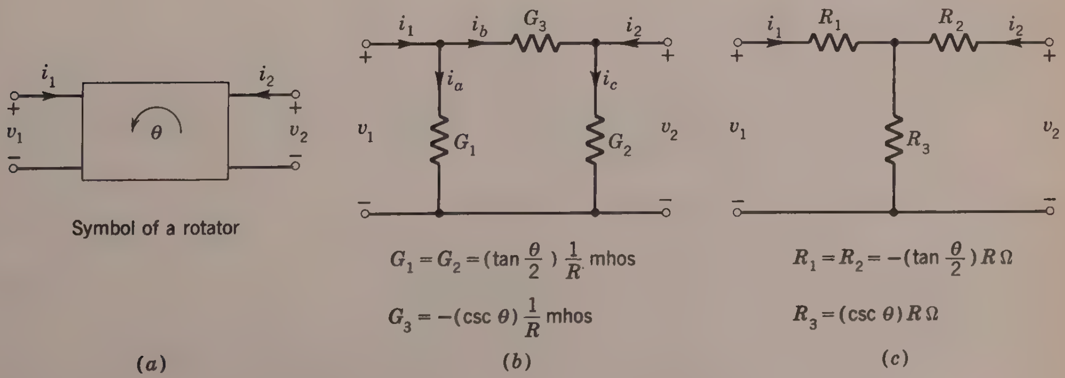
Equations (3-90) and (3-91) can be simplified to

$$i_1 = -(\cot\theta)\frac{1}{R}v_1 + (\csc\theta)\frac{1}{R}v_2 \tag{3-92}$$

$$i_2 = (\csc\theta)\frac{1}{R}v_1 - (\cot\theta)\frac{1}{R}v_2 \tag{3-93}$$

Solving Eq. (3-93) for v_1, we obtain Eq. (3-88). Substituting this expression for v_1 in Eq. (3-92), we obtain Eq. (3-89). This proves that the π network of Fig. 3-32*a* is indeed a realization for a rotator. The T-network realization in Fig. 3-32*c* can be verified by a similar procedure.

Observe that, depending upon the values of θ, either one or two of the three linear resistors in both realizations may assume negative values. However, only one negative resistor is necessary to realize a rotator with any angle of rotation, provided we choose the π network whenever $0° < \theta < 180°$, and the T network whenever $180° < \theta < 360°$. The negative resistor can be realized (over a limited range of voltages) in practice by various practical circuits. A typical rotator realization using transistors is given in Appendix C. For demonstration purposes, an actual oscilloscope tracing of a typical *v-i* curve is shown in Fig. 3-33, together with various rotated v_1-i_1 curves.



Symbol of a rotator

$G_1 = G_2 = (\tan\frac{\theta}{2})\frac{1}{R}$ mhos

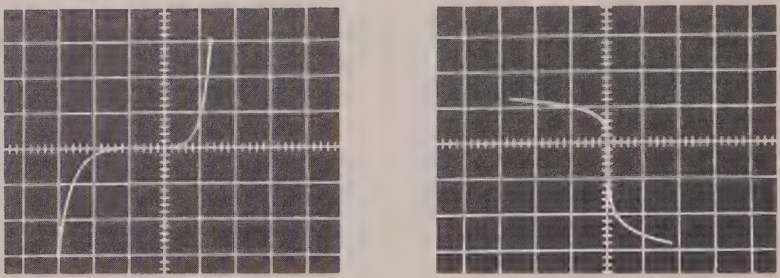
$G_3 = -(\csc\theta)\frac{1}{R}$ mhos

$R_1 = R_2 = -(\tan\frac{\theta}{2})R\ \Omega$

$R_3 = (\csc\theta)R\ \Omega$

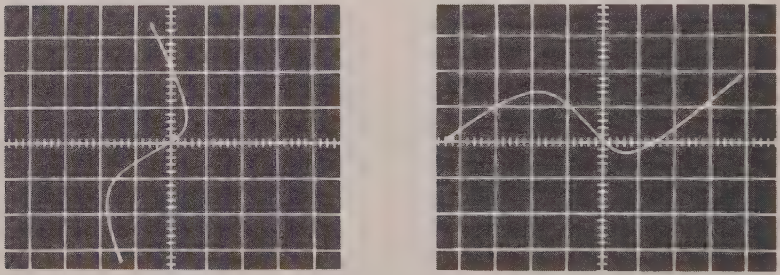
(*a*) (*b*) (*c*)

Fig. 3-32. The rotator, whose symbol is shown in (*a*), can be realized in practice by (*b*) a π network or (*c*) a T network of linear resistors.



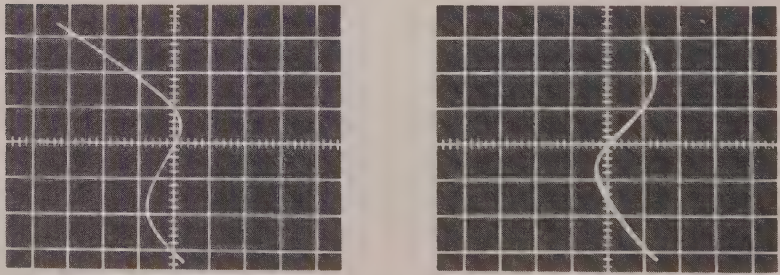
Original v-i curve

v-i curve rotated by 90°



v-i curve rotated by 30°

v-i curve rotated by −45°



v-i curve rotated by 60°

v-i curve rotated by −135°

Fig. 3-33. Actual oscilloscope tracings of a typical v-i curve and the transformed v_1-i_1 curves corresponding to various angles of rotation.

3-7-3 REFLECTORS

By a geometrical procedure similar to that used in deriving the rotator characteristics, we found that a reflector must be characterized by

$$v_1 = (\cos 2\theta)v_2 - (\sin 2\theta)(Ri_2) \tag{3-94}$$

$$i_1 = (\sin 2\theta)\frac{v_2}{R} + (\cos 2\theta)(i_2) \tag{3-95}$$

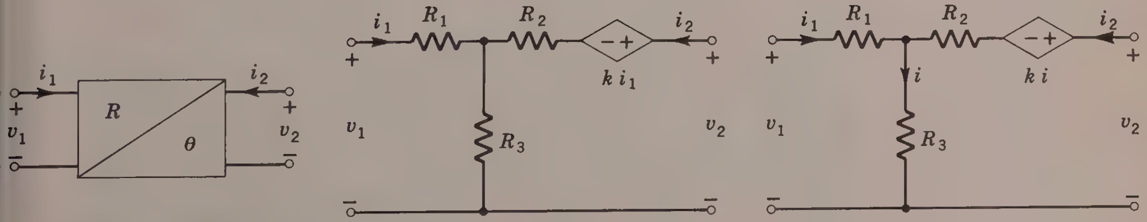
The parameter θ is the angle which the line of reflection (through the origin) makes with the horizontal axis. The constant R is a scale factor having exactly the same significance as for the rotator. For example, if a *v-i* curve is to be reflected in the volt-milliampere plane, then $R = 10^3$. We shall denote a reflector by the symbol shown in Fig. 3-34*a*. By a procedure similar to that used for the rotator, it can be easily shown that a reflector can be realized by the circuits shown in Fig. 3-34*b* and *c*. Observe that the controlled source in Fig. 3-34*b* is controlled by the current i_1 in resistor R_1, whereas the controlled source in Fig. 3-34*c* is controlled by the current i in resistor R_3. Observe also that only one negative resistor is needed if we pick the circuit in Fig. 3-34*b* for $0° < \theta < 90°$ and the circuit in Fig. 3-34*c* for $90° < \theta < 180°$. A practical reflector circuit using transistors is given in Appendix C. For demonstration purposes, several oscilloscope tracings on the reflection of *v-i* curves are shown in Fig. 3-35. The bottom trace is that of an *n-p-n* transistor with its collector and emitter terminals connected across port 2 of a 45° reflector. Each curve in these tracings corresponds to a particular value of base current. It is interesting to observe that this "reflected" transistor can be considered as a new three-terminal element. For this particular example, the almost vertical v_1-i_1 curves can be used to simulate a current-controlled voltage source operating in the first quadrant.

There are several angles of reflection which are of special practical importance, and the corresponding reflector has been given special names as follows:

Gyrator When $\theta = 45°$, Eqs. (3-94) and (3-95) can be recast into the form

$$\begin{aligned} i_1 &= Gv_2 \\ i_2 &= -Gv_1 \end{aligned} \tag{3-96}$$

Fig. 3-34. A reflector can be realized in practice by a controlled source and three linear resistors.



Symbol of a reflector

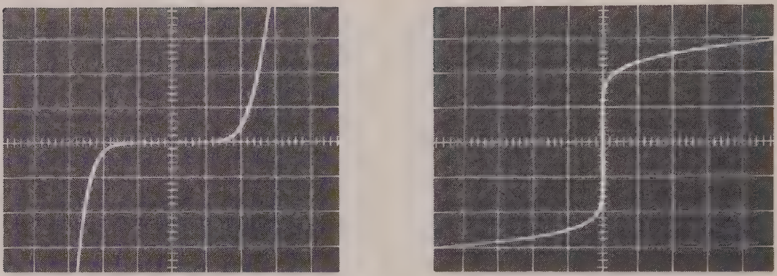
(*a*)

$R_1 = (\cot\theta)\,R\ \Omega$
$R_2 = (\tan\theta)\,R\ \Omega$
$R_3 = -(\csc 2\theta)\,R\ \Omega$
$k = 2\,(\csc 2\theta)\,R$

(*b*)

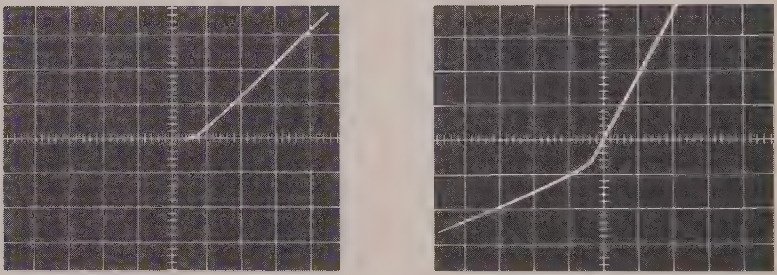
$R_1 = (\cot\theta)\,R\ \Omega$
$R_2 = -(\cot\theta)\,R\ \Omega$
$R_3 = -(\csc 2\theta)\,R\ \Omega$
$k = 2\,(\csc 2\theta)\,R$

(*c*)



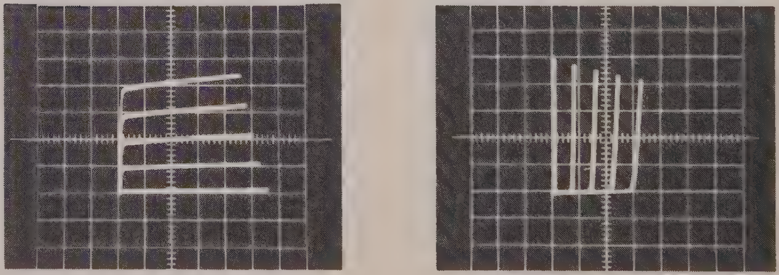
Original v - i curve

v - i curve reflected about the 45° line



Original v - i curve

v - i curve reflected about the 120° line



Original family of v - i curves

Family of v - i curves reflected about the 45° line

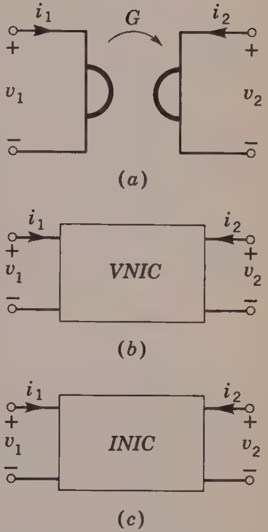
Fig. 3-35. Oscilloscope tracings of typical v-i curves and their reflected v_1-i_1 curves.

where $G \equiv 1/R$. The resulting two-port resistor is called a *gyrator*. The scale factor G in this case is usually called the *gyration conductance*. Observe that a gyrator has the effect of reflecting any v-i curve about the 45° line through the origin. Hence, a voltage-controlled nonlinear resistor becomes a current-controlled resistor and vice versa. The usual symbol for a gyrator is shown in Fig. 3-36*a*.

Voltage-inversion negative-impedance converter (VNIC) When $\theta = 90°$, Eqs. (3-94) and (3-95) become

$$\begin{aligned} v_1 &= -v_2 \\ i_1 &= -i_2 \end{aligned} \tag{3-97}$$

The resulting two-port resistor is called a *voltage-inversion negative-impedance converter,* or simply a VNIC. A VNIC has the effect of reflecting a v-i curve about the vertical axis. Hence, if a linear resistor with positive resistance R is connected across port 2 of a VNIC, the new v_1-i_1 curve becomes a linear resistor with a negative resistance $-R$. This is one simple method for obtaining a negative resistance in practice. We shall denote a VNIC by the symbol shown in Fig. 3-36*b*.



(*a*)

(*b*)

(*c*)

Fig. 3-36. The symbols for a gyrator, a VNIC, and an INIC.

Current-inversion negative-impedance converter (INIC) When $\theta = 180°$, Eqs. (3-94) and (3-95) become

$$\begin{aligned} v_1 &= v_2 \\ i_1 &= i_2 \end{aligned} \tag{3-98}$$

The resulting two-port resistor is called a *current-inversion negative-impedance converter,* or simply an INIC. An INIC has the effect of reflecting a v-i curve about the horizontal axis. Hence, it has the same capability as the VNIC for transforming a positive resistance R into a negative resistance $-R$. We shall denote an INIC by the symbol shown in Fig. 3-36*c*.[1]

In addition to making available a much larger class of nonlinear resistors, the scalor, rotator, and reflector play a much more fundamental role. They are building blocks for synthesizing nonlinear resistors with almost any prescribed v-i curve. This property will be explored in detail in Chap. 8.

The interested reader may find it instructive to build a scalor, a rotator, and a reflector from the circuits given in Appendix C. He will observe, among other things, that these circuits may under certain conditions become *unstable;* i.e., the circuit may burst into unwanted oscillations. The concept of stability will be discussed in Chap. 14. It will be shown there that the oscillation comes from the parasitic elements which are invariably present. Although there are means to stop the oscillation, if it occurs, they are beyond the scope of this book.

Exercise 1: (*a*) Find the equations describing a two-port resistor obtained by connecting a voltage scalor in cascade with a current scalor. (*b*) Discuss the geometrical interpretation achieved by this new transformation. (*c*) Can you suggest some practical applications?

---

[1] There are simple circuits using transistors for realizing a VNIC or an INIC. A typical circuit for each is given in Appendix C.

Exercise 2: A $\theta = 90°$ rotator is also known as a *negative-impedance inverter.* (*a*) If a linear resistor with a positive resistance R is connected across this rotator, find the property of the new resistor. (*b*) Show that when two $\theta = 90°$ rotators are connected in cascade, the result is an ideal transformer with $n = -1$.

Exercise 3: Verify that the T network shown in Fig. 3-32*c* realizes a $\theta°$ rotator.

Exercise 4: Show that when a $\theta_1°$ rotator is connected in cascade with a $\theta_2°$ rotator, the result is a $(\theta_1 + \theta_2)°$ rotator. Does the order of the interconnection matter in this case?

Exercise 5: Verify that each of the circuits shown in Fig. 3-34 realizes a $\theta°$ reflector.

Exercise 6: Show that a $\theta_1°$ reflector in cascade with a $\theta_2°$ reflector produces a $\theta°$ rotator. Find θ in terms of θ_1 and θ_2. Does the order of the interconnection matter in this case?

Exercise 7: Show that a $\theta_1°$ rotator in cascade with a $\theta_2°$ reflector produces a $\theta°$ reflector. Find θ in terms of θ_1 and θ_2. Does the order of the interconnection matter in this case?

Exercise 8: Find the property exhibited by a $\theta = 135°$ reflector. Compare this property with that of a gyrator.

3-8 MUTATOR: THE CHAMELEON BLACK BOX

The two-port resistors discussed in the preceding section share the common property that each transforms a nonlinear resistor into another nonlinear resistor. If we liken the three basic network elements, resistors, capacitors, and inductors, to three distinct species in the genetic sense, then the scalor, rotator, and reflector can be said to transform elements belonging to the same species. They stay within the same race, so to speak.

We shall demonstrate shortly that it is possible to produce a *mutation* from one species into another with the help of a two-port black box called the *mutator.* For example, it is possible to connect a resistor across port 2 of a mutator and produce an inductor across port 1. Conversely, if an inductor is connected across port 1 of the same mutator, a resistor is produced across port 2. For this reason, this class of mutators is called *R-L mutators.* The two other classes are the *R-C mutators* and the *C-L mutators.* An *R-C* mutator transforms a resistor into a capacitor, and vice versa. Similarly, a *C-L* mutator transforms a capacitor into an inductor, and vice versa.

3-8-1 *R-L* MUTATORS

In order to transform a resistor into an inductor, it is necessary that the coordinates (v,i) of each point on a v-i curve be transformed into a corresponding point with coordinates (φ,i) or (i,φ), respectively. Assuming the resistor to be connected across port 2, we can define two types of *R-L* mutators:

Type 1 *R-L* mutator In this case, we require $\varphi_1 = v$ and $i_1 = i$. But $v = v_2$, $i = -i_2$, and $v_1 = d\varphi_1/dt$; therefore, the required relationship for a type 1 *R-L* mutator is

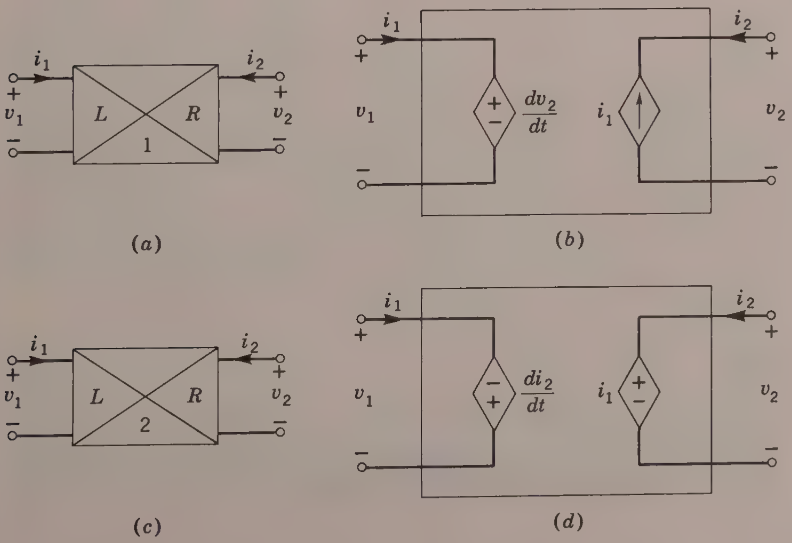
$$\begin{aligned} v_1 &= \frac{dv_2}{dt} \\ i_1 &= -i_2 \end{aligned} \tag{3-99}$$

The symbol for a type 1 *R-L* mutator is shown in Fig. 3-37*a*, and a simple realization using controlled sources is shown in Fig. 3-37*b*.

Type 2 *R-L* mutator In this case, we require $\varphi_1 = i$ and $i_1 = v$. Substituting the relationship $i_2 = -i$, $v_2 = v$, we obtain for a type 2 *R-L* mutator the relationship:

$$\begin{aligned} v_1 &= -\frac{di_2}{dt} \\ i_1 &= v_2 \end{aligned} \tag{3-100}$$

The symbol for a type 2 *R-L* mutator is shown in Fig. 3-37*c*, and a simple realization using controlled sources is shown in Fig. 3-37*d*.



(*a*) (*b*)

(*c*) (*d*)

Fig. 3-37. Two types of *R-L* mutators and a controlled-source realization for each.

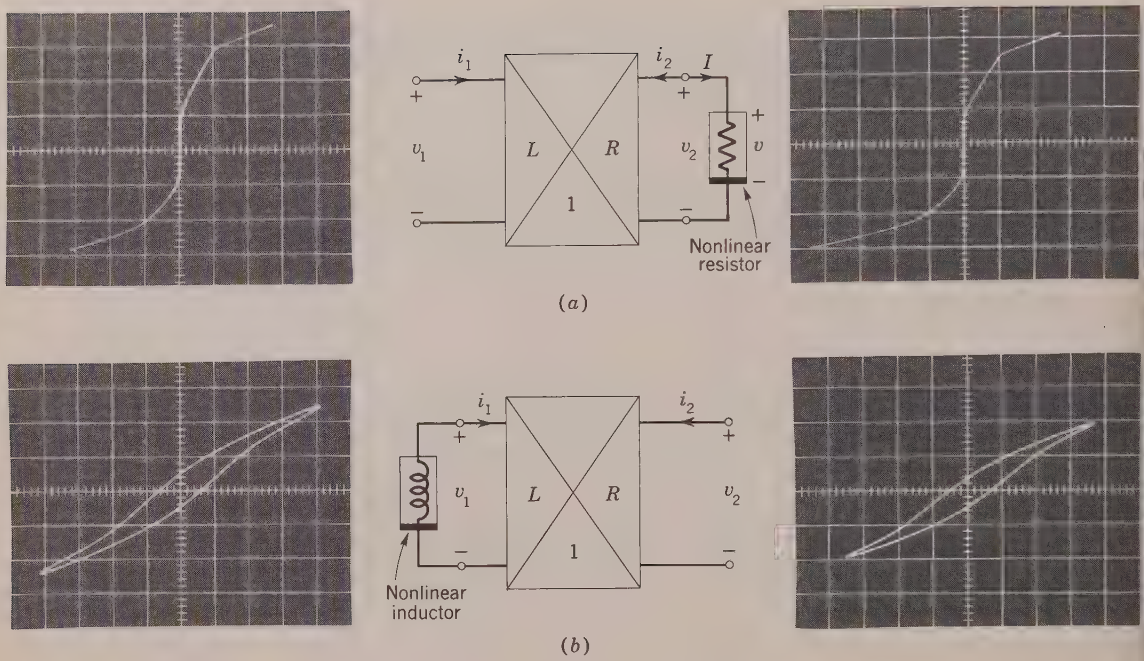


Fig. 3-38. A type 1 *R-L* mutator is used (*a*) to transform a nonlinear resistor into a nonlinear inductor having the same characteristic curve but different terminal variables and (*b*) to transform a nonlinear inductor into a nonlinear resistor having an identical hysteresis curve.

Both types of *R-L* mutators can be realized by circuits using transistors and operational amplifiers. A typical circuit for a type 1 *R-L* mutator is given in Appendix C. To demonstrate the property of an *R-L* mutator, two oscilloscope tracings are shown in Fig. 3-38*a*. The one on the right represents the v-i curve (with voltage v as the vertical axis) of a typical nonlinear resistor connected across port 2 of a type 1 *R-L* mutator. The resulting i-φ curve measured across port 1 is shown on the left (with the flux linkage φ as the vertical axis). To demonstrate that the mutator works in both directions, a typical iron-core inductor is connected across port 1 of *the same* mutator as shown in Fig. 3-38*b*. The oscilloscope tracing on the left shows the hysteresis i-φ curve (with φ as the vertical axis) of the inductor. The "almost" identical curve shown on the right is the v_2-i_2 curve (with the voltage v as the vertical axis) measured across port 2 of the mutator.[1]

3-8-2 *R-C* MUTATORS

To transform a resistor into a capacitor, we may either transform each point on the v-i curve into a corresponding point on the v-q curve, or we may transform each point on the v-i curve into

---

[1] We have departed from our usual practice by plotting the v-i curve with voltage as the vertical axis in order to emphasize that the same curve is being transformed.

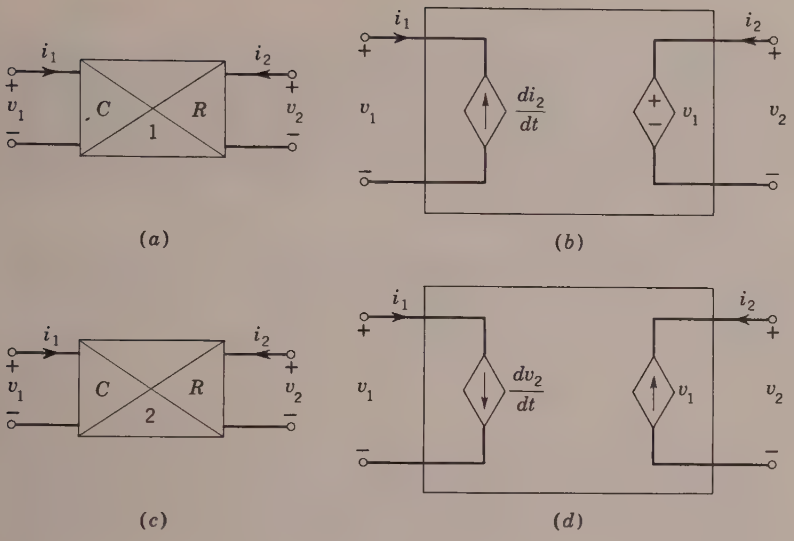


Fig. 3-39. Two types of R-C mutators and a controlled-source realization for each.

a corresponding point on the q-v curve. Accordingly, we can define two types of R-C mutators (assuming that the resistor is connected across port 2):

Type 1 R-C mutator In this case, we require $v_1 = v$ and $q_1 = i$. Since $v_2 = v$ and $i_2 = -i$, a type 1 R-C mutator is characterized by

$$\begin{aligned} v_1 &= v_2 \\ i_1 &= -\frac{di_2}{dt} \end{aligned} \tag{3-101}$$

The symbol for a type 1 R-C mutator is shown in Fig. 3-39*a*, and a simple realization using controlled sources is shown in Fig. 3-39*b*.[1]

Type 2 R-C mutator In this case, we require $v_1 = i$ and $q_1 = v$. Hence, a type 2 R-C mutator is characterized by

$$\begin{aligned} v_1 &= -i_2 \\ i_1 &= \frac{dv_2}{dt} \end{aligned} \tag{3-102}$$

The symbol for a type 2 R-C mutator is shown in Fig. 3-39*c*, and a simple realization using controlled sources is shown in Fig. 3-39*d*.

---

[1] A type 1 R-C mutator circuit using transistors and operational amplifiers is given in Appendix C.

3-8-3 *C-L* MUTATORS

To transform a capacitor into an inductor, we may either transform each point on the v-q curve into a corresponding point on the i-φ curve, or we may transform each point on the v-q curve into a corresponding point on the φ-i curve. Accordingly, we have also two types of *C-L* mutators:

Type 1 *C-L* mutator In this case, we require $i_1 = v$ and $\varphi_1 = q$. Since $v_2 = v$ and $i_2 = -i$, a type 1 *C-L* mutator is characterized by

$$\begin{aligned} v_1 &= -i_2 \\ i_1 &= v_2 \end{aligned} \tag{3-103}$$

The symbol for a type 1 *C-L* mutator is shown in Fig. 3-40*a*, and a simple realization using controlled sources is shown in Fig. 3-40*b*. Comparing Eq. (3-103) with Eq. (3-96), we observe that a type 1 *C-L* mutator is identical with a gyrator with a unity gyration conductance.

Type 2 *C-L* mutator In this case, we require $\varphi_1 = v$ and $i_1 = q$. Hence, a type 2 *C-L* mutator is characterized by

$$\begin{aligned} v_1 &= \frac{dv_2}{dt} \\ i_1 &= -\int_{-\infty}^{t} i_2(\tau)\, d\tau \end{aligned} \tag{3-104}$$

The symbol for a type 2 *C-L* mutator is shown in Fig. 3-40*c*, and a simple realization using controlled sources is shown in Fig. 3-40*d*.

The significance of mutators is that they make it possible to synthesize a nonlinear inductor with a prescribed i-φ curve or a nonlinear capacitor with a prescribed v-q curve, provided we know how to synthesize a nonlinear resistor having an identical v-i curve. In other words, the problem of realizing a nonlinear inductor or a nonlinear capacitor reduces to that of realizing a nonlinear resistor, a subject to be considered in detail in Chap. 8.

Another area where a mutator could be useful is that of integrated circuits, where the present-day inductors are incompatible because of their prohibitive size. It is possible to simulate a linear inductor by connecting a linear resistor across port 2 of an *R-L* mutator or by connecting a linear capacitor across port 2 of a *C-L* mutator, also known as a gyrator. The above approach

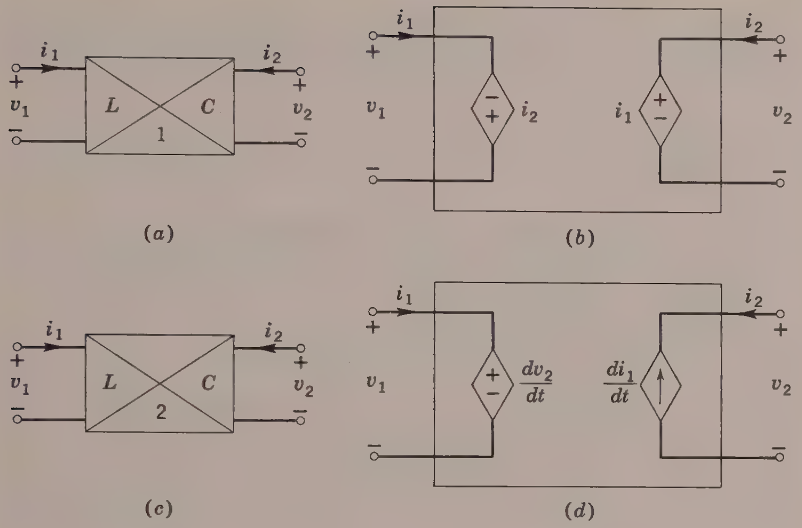


Fig. 3-40. Two types of *C-L* mutators and a controlled-source realization for each.

is still in the research stage because of two basic practical problems. First, due to the presence of parasitic elements, the mutators can not yet be operated economically at very high frequencies. Secondly, just as for scalors, rotators, and reflectors, a mutator, and for that matter, any circuit using transistors and operational amplifiers, may become unstable under certain conditions.

Exercise 1: A type 1 *R-L* mutator can also be realized by a controlled voltage source with a terminal voltage equal to either $(dv_2/dt - v_2)$ or $\left(\int_{-\infty}^{t} v_1(\tau)\,d\tau - v_1\right)$. Find an appropriate circuit realization for each.

Exercise 2: A type 2 *R-L* mutator can also be realized by two controlled current sources with terminal currents equal to v_2 and $\int_{-\infty}^{t} v_1(\tau)\,d\tau$, respectively. Find an appropriate circuit realization.

Exercise 3: A type 1 *R-C* mutator can also be realized by a controlled current source with a terminal current equal to either $(i_2 - di_2/dt)$ or $\left(i_1 - \int_{-\infty}^{t} i_1(\tau)\,d\tau\right)$. Find an appropriate circuit realization for each.

Exercise 4: A type 2 *R-C* mutator can also be realized by two controlled voltage sources with terminal voltages equal to i_2 and $\int_{-\infty}^{t} i_1(\tau)\,d\tau$, respectively. Find an appropriate circuit realization.

Exercise 5: Find another circuit realization for a type 1 *C-L* mutator using two voltage-controlled current sources.

Exercise 6: Find another circuit realization for a type 2 C-L mutator using a controlled current source across port 1 and a controlled voltage source across port 2.

3-9 SUMMARY

Three basic classifications of multiterminal elements

1. N-terminal resistor involving only terminal voltages v_j and terminal currents i_j, $j = 1, 2, \ldots, n$
2. N-terminal capacitor involving only terminal voltages v_j and terminal charges q_j, $j = 1, 2, \ldots, n$
3. N-terminal inductor involving only terminal currents i_j and terminal flux linkages φ_j, $j = 1, 2, \ldots, n$

A three-terminal resistor, capacitor, or inductor is characterized by two sets of curves measured with respect to an arbitrarily chosen ground terminal.

Forms of representation The characteristic curves of a three-terminal resistor can be completely specified by at least six independent representations:

Conductance representation

$$i_1 = i_1(v_1,v_2)$$
$$i_2 = i_2(v_2,v_1)$$

Resistance representation

$$v_1 = v_1(i_1,i_2)$$
$$v_2 = v_2(i_2,i_1)$$

Conductance-resistance representation

$$i_1 = i_1(v_1,i_2)$$
$$v_2 = v_2(i_2,v_1)$$

Resistance-conductance representation

$$v_1 = v_1(i_1,v_2)$$
$$i_2 = i_2(v_2,i_1)$$

Voltage ratio–current ratio representation

$$v_1 = v_1(v_2,i_2)$$
$$i_1 = i_1(v_2,i_2)$$

Current ratio–voltage ratio representation

$$v_2 = v_2(v_1,i_1)$$
$$i_2 = i_2(v_1,i_1)$$

The characteristic curves of a three-terminal resistor with respect to a given ground terminal can always be transformed into an equivalent set of curves with respect to another ground terminal.

It is generally impractical to completely characterize a multiterminal element because a prohibitive number of measurements may be required. To overcome this difficulty, a multiterminal resistor is usually operated in one of the following ways:

1. As a three-terminal controlled resistor: all but three terminals are connected to independent voltage sources or independent current sources.
2. As an element with prescribed constraints
   a. Internal current and voltage constraints. Examples: operational amplifier, AND gate, and OR gate.
   b. External port constraints: current entering one terminal of each port is constrained to leave the other terminal of the same port. A port may be created across any pair of terminals of a multiterminal black box by connecting the terminal pair across an isolated two-terminal black box or across a two-port ideal transformer.

Practical applications of operational amplifiers

1. Nonlinear mode: square-wave converter, voltage comparator, or threshold detectors.
2. Linear mode: scaling circuit, isolation box, phase inverter, differentiator, integrator, and summing circuit. These functional circuits constitute the building blocks of analog computers.

An important class of two-port resistors

1. The scalor multiplies the abscissa or the ordinate of a *v*-*i* curve by a constant.

   a. Voltage scalor:

   $$v_1 = k_v v_2$$
   $$i = -i_2$$

   b. Current scalor:

   $$v_1 = v_2$$
   $$i_1 = -k_i i_2$$

2. The rotator rotates a *v*-*i* curve by a prescribed angle θ with respect to the origin.

   $$v_1 = (\cos\theta)v_2 + (\sin\theta)(Ri_2)$$
   $$i_1 = (\sin\theta)(v_2/R) - (\cos\theta)i_2$$

3. The reflector reflects a *v*-*i* curve with respect to a straight line making an angle θ with the horizontal axis.

   $$v_1 = (\cos 2\theta)v_2 - (\sin 2\theta)(Ri_2)$$
   $$i_1 = (\sin 2\theta)(v_2/R) + (\cos 2\theta)i_2$$

Mutator, the chameleon black box

1. The *R-L* mutator transforms a resistor into an inductor, and vice versa.

Type 1. $(v,i) \leftrightarrow (\varphi_1,i_1)$

$$v_1 = \frac{dv_2}{dt}$$

$$i_1 = -i_2$$

Type 2. $(v,i) \leftrightarrow (i_1,\varphi_1)$

$$v_1 = -\frac{di_2}{dt}$$

$$i_1 = v_2$$

2. The *R-C* mutator transforms a resistor into a capacitor, and vice versa.

Type 1. $(v,i) \leftrightarrow (v,q)$

$$v_1 = v_2$$

$$i_1 = -\frac{di_2}{dt}$$

Type 2. $(v,i) \leftrightarrow (q,v)$

$$v_1 = -i_2$$

$$i_1 = \frac{dv_2}{dt}$$

3. The *C-L* mutator transforms a capacitor into an inductor, and vice versa.

Type 1. $(v,q) \leftrightarrow (i,\varphi)$

$$v_1 = -i_2$$

$$i_1 = v_2$$

Type 2. $(v,q) \leftrightarrow (\varphi,i)$

$$v_1 = \frac{dv_2}{dt}$$

$$i_1 = \int_{-\infty}^{t} i_2(\tau)\,d\tau$$

PROBLEMS

3-1 A three-terminal resistor is characterized by the equations

$$i_1 + v_2 - v_1^3 = 0$$

$$i_2 + v_1 = 0$$

(a) Recast these equations into the six forms of representation given in the text. If the equations cannot be expressed in a particular form, we say the corresponding representation does not exist. Under what conditions may this situation occur?

(b) Sketch the characteristic curves corresponding to the representations obtained in (*a*).

(c) Based on the complexity of the curves obtained in (*a*), explain why a certain form of representation may be preferred over another in practice?

(d) Obtain the characteristic curves in the resistance representation *graphically* from the curves in the conductance representation in (*b*).

3-2 The characteristic curves of most transistors are measured with respect to the common-emitter configuration shown in Fig. P3-2*a*.

From the characteristics of the *n-p-n* transistor (type 2NXX17A) given in Appendix D, obtain the corresponding sets of curves for the common-collector configuration shown in Fig. P3-2*b*.

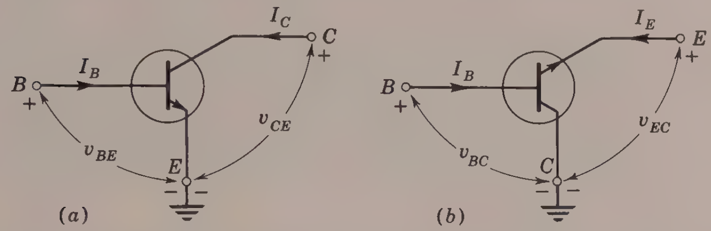


Fig. P3-2.

3-3 The common-base characteristic curves of an *n-p-n* or *p-n-p* transistor can be represented approximately by the so-called Ebers-Moll equations,

$$I_E = A_{11}(e^{kV_{EB}} - 1) + A_{12}(e^{kV_{CB}} - 1)$$

$$I_C = A_{21}(e^{kV_{EB}} - 1) + A_{22}(e^{kV_{CB}} - 1)$$

where k is a constant and A_{11}, A_{12}, A_{21}, and A_{22} are coefficients which depend upon the type and construction of the transistor.

(a) Find the corresponding equations in the common-emitter configuration.

(b) Repeat (*a*) for the common-collector configuration.

(c) Find the appropriate sign of the coefficients A_{11}, A_{12}, A_{21}, and A_{22} for an *n-p-n* transistor.

(d) Repeat (*c*) for a *p-n-p* transistor.

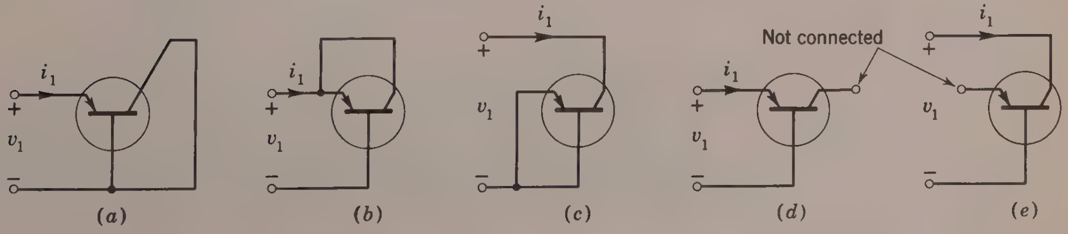
(e) Discuss any similarities or differences between the Ebers-Moll equations and Eq. (1-14) of a junction diode. Can you find a simple relationship between them?

3-4 In the fabrication of large-scale integrated circuits, it is highly economical to produce many identical transistors and then to obtain from them other elements such as diodes. In the case of diodes, the transistor is operated as a two-terminal controlled resistor and may take the five basic forms shown in Fig. P3-4.

(a) Obtain the v_1-i_1 curve of each of these five configurations. Assume a *p-n-p* transistor type 2NXX16.

(b) How should you redefine the references if the *p-n-p* transistor is replaced by an *n-p-n* transistor?

Fig. P3-4.



3-5 There is an important class of multiterminal black boxes, designed specifically for processing digital signals which can assume only one of two distinct levels, namely, $v_i = E_0$ or $v_i = E_1$. See Fig. P3-5*a*. It is traditional to attach a binary code of 0 or 1 to represent these two distinct states, respectively. In order to study the various possible state combinations, it is convenient to tabulate the state of the output voltage corresponding to the state of each input voltage as shown in Fig. P3-5*b*. This table is usually called a *truth table*. Construct the truth table to represent the following common logic circuits:

(a) An AND gate.

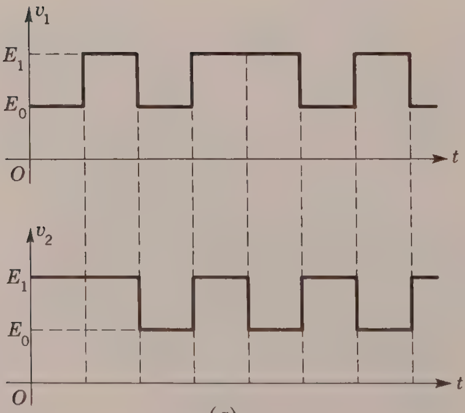
(b) An OR gate.

(c) A black box characterized by $v_o = E_1$, whenever $v_1 = v_2 = E_0$. Otherwise, $v_o = E_0$. This logic circuit is usually called an *INVERTING gate* or a *NOT gate*.

(d) A black box characterized by $v_o = E_1$ whenever $v_1 = E_1$ *or* $v_2 = E_1$, but not both. Otherwise, $v_o = E_0$. This logic circuit is usually called an *exclusive OR gate*.

(e) A black box characterized by $v_o = E_0$ whenever $v_1 = v_2 = E_1$. Otherwise, $v_o = E_0$. This logic circuit is usually called a *NAND gate*.

(f) A black box characterized by $v_o = E_1$ whenever $v_1 = v_2 = E_0$. Otherwise, $v_o = E_0$. This logic circuit is usually called a *NOR gate*.



Input signal		Output signal
v_1	v_2	v_o
0	0	
0	1	
1	0	
0	0	

Fig. P3-5. (*a*) (*b*)

3-6 A two-port ideal transformer can be approximately realized in practice by a suitable interconnection of a voltage-controlled voltage source and a current-controlled current source. Draw the appropriate circuit and specify the relationship between the controlling coefficients so as to obtain a turns ratio equal to n_1/n_2.

3-7 The equation of a rocket in motion can be approximately represented by the following differential equation:

$$a_4\frac{d^4x}{dt^4} - a_3\frac{d^3x}{dt^3} + a_2\frac{d^2x}{dt^2} + a_1\frac{dx}{dt} - a_0 = f(t)$$

where a_0, a_1, a_2, a_3, and a_4 are positive coefficients which can be adjusted by the ground control to keep the rocket in proper operation. The problem is to study the effect of these parameters on the performance of the rocket. Although an actual rocket could be launched for this purpose, it is extremely costly to do so. Show how you may obtain the information by an appropriate analog circuit.

3-8 The analog circuits discussed in the text generally require two operational amplifiers to perform pure "subtraction" or pure "addition." The following circuits require only one operational amplifier for the same task.

(a) Show that the circuit in Fig. P3-8*a* is characterized by $v_o = k(v_2 - v_1)$, where $k = R_0/R_1$. This circuit is also known as a *differential input amplifier.* Explain why.

(b) Show that the circuit in Fig. P3-8*b* is characterized by $v_o = v_1 + v_2$, provided $R_0 = 2R_1$.

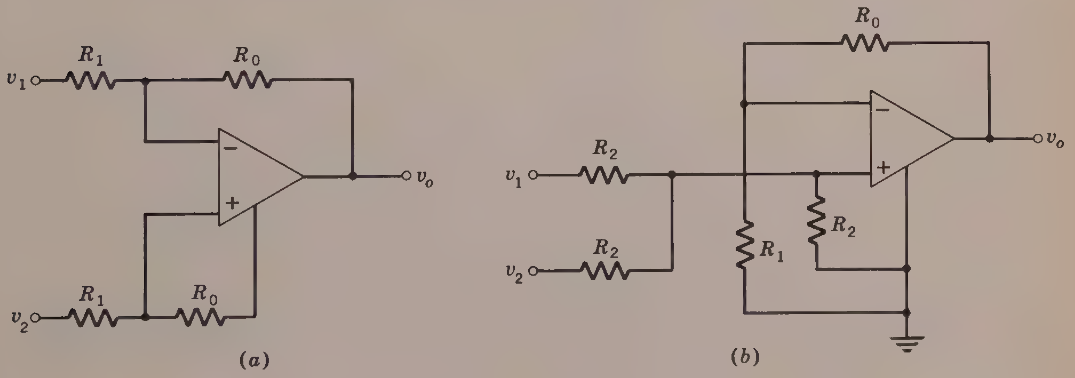


Fig. P3-8.

3-9 (a) Show that when an inductor is connected across port 2 of a voltage scalor, the resulting two-terminal black box is also an inductor having the same *i*-*φ* curve, but with its flux linkage multiplied by the constant k_v.

(b) For a current scalor, show that the same *i*-*φ* curve is obtained, but with the current multiplied by the constant k_i.

(c) Show that an analogous property holds for a capacitor.

3-10 (a) Show that an analogous two-port black box called an *L* rotator may be defined to rotate the *i*-*φ* curve of a nonlinear inductor.

(b) Show that an L rotator can be realized in practice by either a π network or a T network containing three linear inductors.

(c) Show that an L rotator can also be realized by connecting either port of a rotator across port R of an R-L mutator.

3-11 (a) Show that an analogous two-port black box called a C rotator may be defined to rotate the v-q curve of a nonlinear capacitor.

(b) Show that a C rotator can be realized in practice by either a π network or a T network containing three linear capacitors.

(c) Show that a C rotator can also be realized by connecting either port of a rotator across port R of an R-C mutator.

3-12 (a) Show that an analogous two-port black box called an L reflector may be defined to reflect the i-φ curve of a nonlinear inductor.

(b) Show that an L reflector can be realized by the same reflector circuits given in the text, provided the resistors are replaced by inductors.

(c) Show that an L reflector can also be realized by connecting either port of a reflector across port R of an R-L mutator.

3-13 (a) Show that an analogous two-port black box called a C reflector may be defined to reflect the v-q curve of a nonlinear capacitor.

(b) Show that a C reflector can be realized by the same reflector circuits given in the text, provided the resistors are replaced by capacitors.

(c) Show that a C reflector can also be realized by connecting either port of a reflector across port R of an R-C mutator.

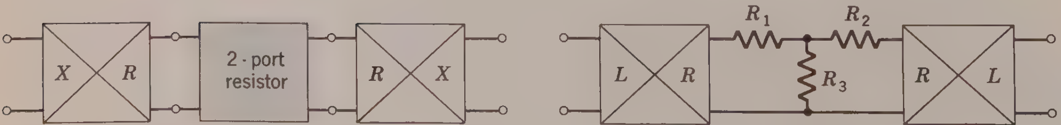
3-14 (a) Show that a two-port resistor may be transformed into a two-port inductor or a two-port capacitor by the circuit shown in Fig. P3-14*a*, where X stands for L in the inductor case and C in the capacitor case.

(b) Show that the circuit shown in Fig. P3-14*b* is equivalent to the three-terminal linear inductor defined by Eqs. (3-41) and (3-42).

(c) In practice, this three-terminal inductor is called a *pair of mutual inductance* if $L_{12} = L_{21} \equiv M$. Find the relationships between L_{11}, L_{22}, M and R_1, R_2, R_3.

(d) Based on the observation that an R-L mutator can be synthesized without using any inductor, discuss the possible application of this mutual inductance in integrated circuits.

Fig. P3-14.



(*a*) (*b*)

3-15 Verify that the two-port black box N_3 is obtained from the cascade connection between N_1 and N_2 as shown in Fig. P3-15.

(a) N_1 = type 1 C-L mutator, N_2 = type 2 R-C mutator, N_3 = type 1 R-L mutator.

(b) N_1 = type 1 C-L mutator, N_2 = type 1 R-C mutator, N_3 = type 2 R-L mutator.

(c) N_1 = type 1 C-L mutator, N_2 = type 2 R-L mutator, N_3 = type 1 R-C mutator.

(d) N_1 = type 1 C-L mutator, N_2 = type 1 R-L mutator, N_3 = type 2 R-C mutator.

(e) N_1 = voltage scalor with $k_v = n$, N_2 = current scalor with $k_i = 1/n$, N_3 = two-port ideal transformer.

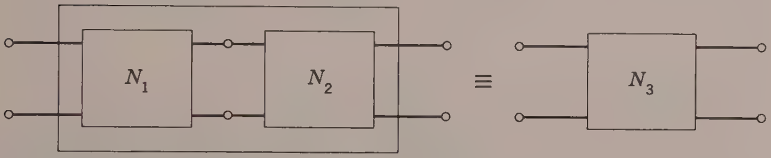


Fig. P3-15.

3-16 The rotator, the reflector, and the scalor can be considered as special cases of a more general two-port resistor called the *linear transformation converter* (LTC). The LTC is characterized by

$$v_1 = Av_2 - Bi_2$$

$$i_1 = Cv_2 - Di_2$$

$$AD - BC \neq 0$$

(a) Find the coefficients A, B, C, and D for the rotator, the reflector, and the scalor.

(b) Prove that any prescribed LTC can be synthesized by connecting in cascade a θ_1° reflector, a current scalor with coefficient K_I, a voltage scalor with coefficient K_V, and a θ_2° rotator, *in the above order.* Find the relationship of θ_1, θ_2, K_I, and K_V in terms of the prescribed coefficients A, B, C, and D.

3-17 A *multiplier* is a four-terminal element (consisting of two input terminals, one output terminal, and a common ground terminal) characterized by $i_1 = 0$, $i_2 = 0$, and $v_o = v_1 v_2$.

(a) Show that a *squaring circuit* characterized by $i_1 = 0$, $i_2 = 0$, and $v_o = v_i^2$ can be synthesized using only one multiplier.

(b) Using the squaring circuit from (*a*) and an integrator, synthesize a *mean-squaring circuit* characterized by $i_1 = 0$, $i_2 = 0$, and $v_o(t) = \int_{-\infty}^{t} v_i^2(\tau)\, d\tau$.

(c) Show that a circuit for finding the *square root* of an input signal ($v_o = \sqrt{v_i}$) can be synthesized using only two linear resistors, an operational amplifier, and a multiplier.

(d) Show that a *divider circuit* characterized by $i_1 = 0$, $i_2 = 0$, and $v_o = v_1/v_2$ can be synthesized using only two linear resistors, an operational amplifier, and a multiplier.

4 EQUATIONS OF MOTION

4-1 CLASSIFICATION OF NONLINEAR NETWORKS

So far, we have been concerned primarily with the classification of n-terminal black boxes where $n \geq 2$. Regardless of the number of terminals n, we have found it convenient to define only three basic types of lumped network elements, namely, n-terminal resistors, n-terminal capacitors, and n-terminal inductors.[1] The main reason for this three-way classification is that most practical devices either belong to one of these types or may be modeled by an appropriate combination of these elements. We have also discussed various practical applications of each type of elements. Since little background is assumed of the reader at this stage, only simple applications were discussed. Even then, we have found it necessary in many occasions to somewhat idealize the circuits. The reader will discover many more practical applications by the time he finishes this book. It is hoped that these exciting applications will provide the reader with enough motivation to embark on a serious study of nonlinear network theory. In the hope of providing sufficient generality to the principles and methods, we have found it necessary at times to inject a certain amount of formalities which may not be as motivating as concrete, but less general, examples.[2]

Our first formal task is to state a number of important definitions.

DEFINITION OF NETWORKS AND SUBNETWORKS

A network is any interconnection of the three basic types of network elements and sources, which may be both independent and controlled. A network is said to be *linear* if it contains only linear elements (in the sense defined previously) and sources. Otherwise, it is *nonlinear.*[1] On many occasions we shall find it useful to partition the network elements in a network N into two or more groups, where the elements in one group may be connected to elements in another group only through external wires. That part of the network containing any particular group of elements is called a *subnetwork.*

---

[1] To these we may add the class of *independent sources,* which may be interpreted as a special case of two-terminal resistors, and the class of *controlled sources,* which may be interpreted as a special case of four-terminal, or two-port, resistors. There are occasions where it is more convenient to regard these sources as a separate set of elements. We shall find both interpretations useful in this book.

[2] This is necessary in order that the reader may know exactly the conditions under which a particular method is valid. By so doing, the reader will be in a better position to solve new problems that may confront him in the future.

DEFINITION OF RESISTIVE AND DYNAMIC NETWORKS

A network is said to be a *resistive network* if it contains *only* n-terminal resistors and sources.[2] Otherwise, it is called a *dynamic* network.[3]

The reason for the above classification is a very fundamental one. As will be demonstrated shortly, the governing equations describing any resistive nonlinear network are nonlinear functional (algebraic or transcendental) equations. In contrast with this, the equations describing any dynamic nonlinear network are nonlinear differential equations. Since these two types of equations are solved by entirely different techniques, it is only logical to analyze these types of networks separately. Since resistive networks are more fundamental than dynamic networks, we shall study this class of networks first. The importance of resistive-network analysis as a prerequisite to the analysis of dynamic networks can be seen by noting that all part 2 (Chaps. 5 to 12) of this book is addressed only to resistive networks. Although certain simple dynamic networks could have been discussed earlier, or even mixed in with the discussion on resistive networks, we have found it more satisfactory from a pedagogical point of view to postpone all discussion of dynamic networks until part 3.

Our main purpose in this chapter is to formulate a systematic method for writing the governing equations describing a nonlinear network. In analogy to mechanics, we shall henceforth refer to the governing equations as the *equations of motion.*

4-2 LAWS OF ELEMENTS AND OF INTERCONNECTION

The three basic network elements are completely specified by a curve, or sets of curves, relating the variables voltage and current for a resistor, voltage and charge for a capacitor, and current and flux linkage for an inductor. We can interpret the element-characteristic curves as constraints or "laws" that must be satisfied by the element at all times. Hence, we shall refer to the element-characteristic curves as the *laws of elements.*

---

[1] Since almost all networks to be considered in this book are nonlinear, we shall usually omit writing this adjective.

[2] Recall that the class of n-terminal resistors includes a large variety of special cases such as controlled resistors and n-port resistors. In particular, a resistive network may contain ideal transformers, scalors, rotators, reflectors, gyrators, and negative impedance converters.

[3] For reasons that will soon be obvious, a resistive network is sometimes called a *memoryless* network or an *instantaneous* network. In contrast with this, a dynamic network is often called a network with *memory.*

The laws of elements pertain only to individual elements. When these elements are interconnected to form a network, there are some governing laws that all elements in the network must obey. We shall refer to these laws as the *laws of interconnection.* In lumped-network theory, the laws of interconnection consist of the following:

1. Kirchhoff voltage law (KVL): The algebraic sum of voltages around any closed loop of a lumped electrical network is zero at all times.
2. Kirchhoff current law (KCL): The algebraic sum of currents entering any node of a lumped electrical network is zero at all times.

It must be emphasized that the laws of elements and the laws of interconnection are distinct and completely independent laws. Perhaps a simple example from freshman physics will help to make this distinction clear. Consider a mechanical system consisting of a stiff spring coil *A* with one end fixed against the ceiling and the other tied to one end of a smaller spring coil *B*. A mass *M* is attached to the other end of spring *B*. In order to predict the motion when the mass *M* is given an initial displacement and then released, it is necessary to write the equations of motion describing the system. In order to do this, we must know first the characteristics of each of the two springs, which in this case are given by the well-known Hooke's law. But Hooke's law pertains only to each spring and is independent of the entire system. In other words, regardless of how the springs are interconnected, each spring must independently satisfy Hooke's law at all times. In order to complete the equations of motion describing the system, we need some other laws which govern the interconnection of mechanical elements, namely, Newton's laws of mechanics. It is clear, therefore, that for our example, Hooke's law corresponds to the law of elements and Newton's laws correspond to the law of interconnection. Both are needed to write the complete equations of motion of the system.

The fundamental problem in nonlinear network theory is to determine the voltages across all network elements and the currents through all network elements which simultaneously satisfy the laws of elements and the laws of interconnection. We shall call such a set of voltages and currents the *solution of the network.*

When the laws of elements and the laws of interconnection are applied to a given network, the resulting equations are called the *equations of motion.*

4-2-1 EQUATIONS OF MOTION PERTAINING TO THE LAWS OF ELEMENTS

The general form of equations describing the laws of elements has already been discussed in the preceding chapters. For nonlinear elements these equations are also nonlinear and are, therefore, difficult to describe analytically. Nevertheless, it is useful to denote these relationships symbolically by functional notations. For example, the laws describing a two-terminal resistor, capacitor, and inductor can be represented in the following general forms:[1]

$$\text{Resistor:} \quad f_R(v_j,i_j) = 0 \tag{4-1}$$

$$\text{Capacitor:} \quad f_C(v_j,q_j) = 0 \qquad i_j = \frac{dq_j}{dt} \tag{4-2}$$

$$\text{Inductor:} \quad f_L(i_j,\varphi_j) = 0 \qquad v_j = \frac{d\varphi_j}{dt} \tag{4-3}$$

It is important to observe that the variables in each of the above equations pertain only to a particular element. Since each equation involves a different set of variables, it is impossible to derive one equation from the remaining equations. Mathematically, this statement is equivalent to saying that the equations of motion pertaining to the laws of elements are always independent of one another. Another important observation is that whereas the laws of resistors are always functional equations,[2] the laws of capacitors and inductors involve time derivatives of q_j and φ_j and hence are functional-differential equations; i.e., each element must be described by both functional equations and differential equations.

4-2-2 EQUATIONS OF MOTION PERTAINING TO THE LAWS OF INTERCONNECTION

It is clear from the statements of KVL and KCL that the equations arising from the laws of interconnection are independent of the type of elements in the network. In other words, it does not matter whether a certain element across a pair of nodes is a resistor, a capacitor, an inductor, or a source. Only the network connection diagram, *the topology,* need be specified in order to ob-

---

[1] The same general form applies to n-terminal elements, where $n > 2$. The only difference here is that each element is now described by $(n - 1)$ equations. Each equation is generally a function of $2(n - 1)$ variables, namely, v_j, i_j for resistors, v_j, q_j for capacitors, and i_j, φ_j for inductors, with $j = 1, 2, \ldots, n - 1$.

[2] Any operation which does not contain time-derivative or integral terms is said to be a functional equation. In particular, a functional equation is said to be an algebraic equation if it is a polynomial equation. Otherwise, it is a transcendental equation.

tain the equations corresponding to the laws of interconnection. This observation is particularly useful when the network contains only two-terminal elements and sources. Here it is convenient to replace each two-terminal element (Fig. 4-1*a*) in the network by a short line segment called a "branch" or an "edge" of the element as shown in Fig. 4-1*b*. As usual, each terminal of the branch is called a node. Observe that each branch must indicate the original reference direction for i_j and the original reference polarity for v_j. It is common practice, however, to draw only an arrow in each branch, as shown in Fig. 4-1*c*, and to follow the convention that the node at the tail end of the arrow is always the positive terminal.

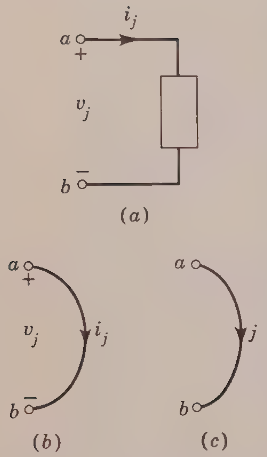


Fig. 4-1. For the purpose of obtaining the equations pertaining to the laws of interconnection, each two-terminal element or source may be replaced by a line segment called a branch. By convention, the positive terminal is assumed to be at the tail end of the branch arrow and need not be indicated explicitly.

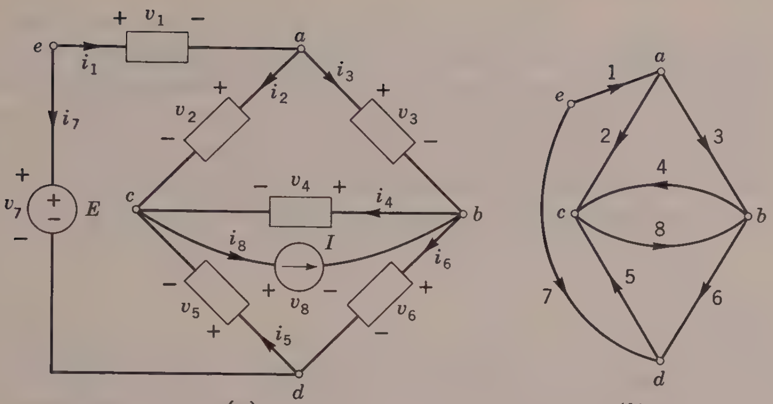
If each two-terminal element of a network N is replaced by a branch, the result is called a *network graph* or simply a graph. To demonstrate that the equations pertaining to the laws of interconnection can be written directly from this graph, let us consider the typical bridge network shown in Fig. 4-2*a* and its network graph as shown in Fig. 4-2*b*. The laws of interconnection can now be obtained as follows:

KVL equations Corresponding to each of the following closed loops, we obtain one equation:

Loop formed by	KVL equation	
Branches 1, 2, 5, 7	$v_1 + v_2 - v_5 - v_7 = 0$	(4-4*a*)
Branches 3, 4, 2	$v_3 + v_4 - v_2 = 0$	(4-4*b*)
Branches 3, 8, 2	$v_3 - v_8 - v_2 = 0$	(4-4*c*)
Branches 4, 8	$v_4 + v_8 = 0$	(4-4*d*)
Branches 4, 6, 5	$-v_4 + v_6 + v_5 = 0$	(4-4*e*)
Branches 8, 6, 5	$v_8 + v_6 + v_5 = 0$	(4-4*f*)
Branches 3, 6, 5, 2	$v_3 + v_6 + v_5 - v_2 = 0$	(4-4*g*)
Branches 1, 3, 6, 7	$v_1 + v_3 + v_6 - v_7 = 0$	(4-4*h*)
Etc.	Etc.	

KCL equations Corresponding to each of the following nodes, we obtain one equation:

Node	KCL equation	
a	$i_1 - i_2 - i_3 = 0$	(4-5*a*)
b	$i_3 - i_4 - i_6 + i_8 = 0$	(4-5*b*)
c	$i_2 + i_4 + i_5 - i_8 = 0$	(4-5*c*)
d	$-i_5 + i_6 + i_7 = 0$	(4-5*d*)
e	$i_1 + i_7 = 0$	(4-5*e*)



(a) (b)

Fig. 4-2. A typical bridge network and its associated network graph.

It should now be clear that the equations arising from the laws of interconnection are always *linear algebraic equations* (whose coefficients are either plus or minus 1) and it is a simple matter to write these equations. However, the trouble with the above equations is that there are too many of them. Many of the equations are redundant in the sense that they can be derived from other equations in the set. For example, Eq. (4-4*d*) can be obtained by subtracting Eq. (4-4*c*) from Eq. (4-4*b*), and therefore it is redundant. Similarly, Eq. (4-5*e*) is also redundant because it can be obtained by adding Eqs. (4-5*a*) to (4-5*d*).

The above examples clearly point out the need to have a systematic procedure for writing an "independent" set of KVL and KCL equations. In order to do this, we must study some elementary concepts of network topology.

Exercise 1: Obtain Eqs. (4-4) and (4-5) directly from the network shown in Fig. 4-2*a*. What is the advantage of using the network graph?

Exercise 2: The *v-i* curves of two nonlinear resistors in a network are given by $v_1 = i_1^3$ and $i_2 = v_2 - v_2^5$. (*a*) Recast these into the form of Eq. (4-1). (*b*) Are these equations independent? Explain why.

Exercise 3: Consider the system of two equations $x_1 + 2x_2 = 4$ and $2x_1 + 4x_2 = 8$. (*a*) Are these equations independent? (*b*) Solve for the value or values of x_1 and x_2. (*c*) How many solutions do these equations possess? From this answer, explain why it is necessary to have independent equations.

4-3 INTRODUCTION TO NETWORK TOPOLOGY

Let *N* be a network containing only two-terminal network elements and independent sources.[1] Let *G* be the network graph associated with *N*. We shall assume that *N* is a *connected network* in the sense that it consists of only one piece; i.e., there is no physical separation. A network graph corresponding to a connected network is said to be a *connected graph.* Clearly, it is possible to move from one node of a connected graph to any other node via a path made up entirely of branches of the graph. Throughout this section, we shall let n denote the total number of nodes and b denote the total number of branches in the network graph G.

---

[1] It is possible to include multiterminal elements as well as controlled sources in a more general discussion of network topology. However, this would involve certain complexities beyond the intended level of this book. When such elements are allowed, the network graph usually consists of several separate subgraphs which are physically disconnected. Because of this, some of the results to be derived in this section may not be applicable in the more general case.

DEFINITION OF TREE AND COTREE

A set of branches of a network graph G is said to be a tree of G if the branches connect all nodes of G but do not form any closed loop. Any branch belonging to a tree T is called a *tree branch.* Branches that do not belong to a given tree are called *links.* All the links of a given tree form what is called a cotree T_c. Therefore, there is a definite cotree associated with every tree.

It is clear from the definition that there are in general many distinct trees for any network. For example, some of the possible trees of the network graph in Fig. 4-2*b* are shown in Fig. 4-3. The cotree corresponding to T_1 then consists of branches 3, 4, 5, and 7. Similarly, the cotree corresponding to the tree T_2 consists of the branches 2, 4, 7, and 8, etc.

Based on the above definition, we can derive several important properties that are valid for any connected network containing only two-terminal elements and independent sources.

PROPERTY 1

Let G be a network graph with n nodes and b branches. Let T be any tree of G and T_c be the corresponding cotree. Then *T contains exactly $(n - 1)$ tree branches and T_c contains exactly $b - (n - 1)$ links.*

In order to prove this property, we must appeal to a method called *mathematical induction.*[1] Since each branch must have at least two nodes, the first step (corresponding to $m = 1$) is to verify that a two-node network graph contains exactly one tree branch. This is obviously the case, for were there two tree branches connecting the two nodes, there would be a loop which is excluded by the definition of a tree. For the second step, let us suppose that it is true that a k-node network graph contains exactly $(k - 1)$ tree branches. If we add one more node to this graph, we must add one more branch to the original tree, for otherwise, no connection would exist from this node to the nodes connecting the tree branches. Again, it is not possible to add more than one tree branch without creating a loop. Hence, our new tree for the $(k+1)$-node network graph has exactly $(k-1)+1=k$ tree branches. This verifies that the first proposition of property 1 is true for all cases. Since G has a total of b branches, the remaining $b-(n-1)$ branches must be links. This proves property 1.

---

[1] According to this method, which has its foundation in mathematical logic, any proposition which claims to be true for all cases $m = 1, 2, \ldots,$ etc. can be proved by verifying two steps: (1) Show that the proposition is indeed true for $m = 1$. (2) Assume next that the proposition is true for $m = k$. Show that this assumption implies that the proposition is also true for $m = k + 1$.

PROPERTY 2

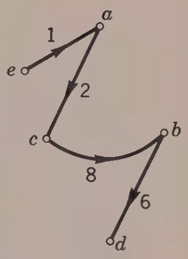
The $n-1$ tree-branch voltages corresponding to any tree of a network without controlled sources are independent of one another; that is, it is impossible to express any tree-branch voltage as a linear combination of the other tree-branch voltages.

This property follows from two observations. (1) Since none of the tree branches is a controlled source, there cannot exist any relationship between two or more tree-branch voltages other than possibly a constraint that may be imposed by KVL. (2) Since by definition the tree branches do not form any closed loop, it is impossible to write a KVL equation that involves only the tree-branch variables. Hence, it is impossible to find any relationship among the tree-branch variables; consequently, it is impossible to solve for any tree-branch voltage in terms of the remaining tree-branch variables.

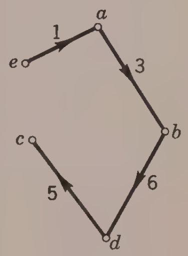
PROPERTY 3

The $b-(n-1)$ link currents corresponding to any cotree of a network without controlled sources are independent of one another.

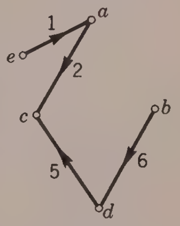
The proof of this property is very similar to that of property 2. First, we observe that since none of the links is a controlled source, there cannot exist any relationship between two or more link currents other than possibly a constraint that may be imposed by KCL. Next, we observe that, by definition, a tree must connect all nodes. This is equivalent to saying that it is impossible to find a node connected only by links, since one of the branches must necessarily be a tree branch. But since KCL pertains to all branches terminating in a node, it is impossible to obtain a KCL equation containing only link currents. Hence, property 3 follows from reasons similar to those given earlier.



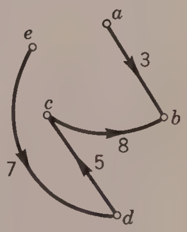
(a) Tree T_1



(b) Tree T_2



(c) Tree T_3



(d) Tree T_4

Fig. 4-3. Four of the many distinct trees of the network graph shown in Fig. 4-2*b*.

Exercise 1: (*a*) Find all trees of the network graph in Fig. 4-2*b* other than those already given in Fig. 4-3. (*b*) List the cotrees corresponding to each tree obtained in (*a*). (*c*) Verify that property 1 is satisfied by each tree obtained in (*a*).

Exercise 2: Verify that properties 2 and 3 are satisfied by the network graph of Fig. 4-2*b*.

4-3-1 CRITERIA FOR WRITING INDEPENDENT KVL EQUATIONS

Our objective in this section is to formulate some criteria for writing a system of independent KVL equations. As usual, we let T be any tree of a network graph G. Since the tree branches connect all nodes but do not form closed loops, it is clear that the addition of any one link will immediately establish a closed loop formed by that link and one or more tree branches. Any such closed loop formed by the tree branches and exactly one link is called a *fundamental loop*.

For example, corresponding to the tree T_1 shown in Fig. 4-3, we obtain one fundamental loop corresponding to each link; hence, corresponding to link 7, we obtain loop 1, 2, 8, 6, 7. Observe that loop 1, 3, 6, 7 is not a fundamental loop because it has more than one link, namely, 3 and 7. Since there are exactly $b - (n - 1)$ links in any network graph in accordance with property 1, it is clear that there are exactly $b - (n - 1)$ fundamental loops in any network graph. We are now ready to state a fundamental theorem which shows us how to write a set of independent KVL equations.

INDEPENDENT KVL EQUATION CRITERIA

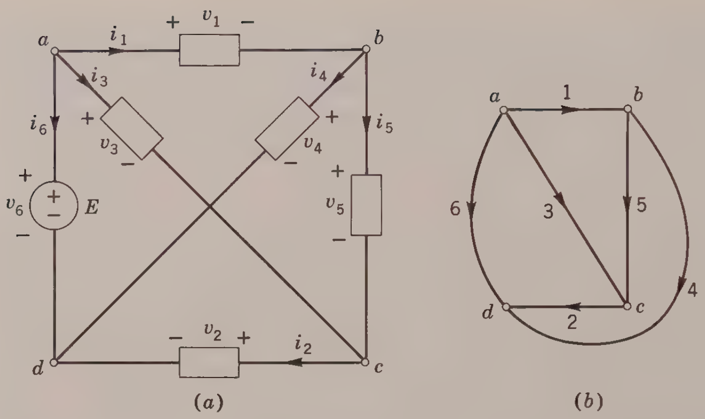
Let G be any network graph containing n nodes and b branches. Let T be any tree in G. Then *the maximum number m of independent KVL equations that can be written is equal to $b - (n - 1)$. Moreover, these m independent equations can always be written by applying KVL to the $b - (n - 1)$ fundamental loops associated with the tree T.*

We shall prove the first assertion by the method of contradiction. Let $m = b - (n - 1)$, and suppose that it is possible to write M independent KVL equations, where $M > m$. Let us consider the first m equations. Clearly they must be independent, too. These m equations would in general contain b unknowns, namely $v_1, v_2, \ldots, v_b$. Since $m < b$, we do not have enough equations to solve for the unknowns. However, in view of property 2 we know the $n - 1$ tree-branch voltages are independent, which means that we can assume any value for these voltages without violating the laws of interconnection. Suppose we let the tree-branch voltages be $v_1, v_2, \ldots, v_{n-1}$ and consider the remaining m link voltages to be the unknowns. Our m independent equations would then contain exactly m unknowns, and we can solve for any of the link voltages uniquely in terms of the $n - 1$ tree-branch voltages. But we have used only m of the M equations. This means that the remaining $M - m$ equations must automatically be satisfied by the solutions we obtained earlier using the m equations, and therefore they can be derived from the m equations. This is a contradiction to our original assumption. Hence, we have shown that there cannot be more than m independent KVL equations.

The above criteria would finally be proved if we show that we can actually write m independent KVL equations. If we choose the loops according to the criteria, then, since each of the $b - (n - 1)$ fundamental loops contains exactly one of the $b - (n - 1)$ links, it is clear that the KVL equation corresponding to each fundamental loop would contain one link voltage which is missing in all other equations. In other words, each of the m equations contains one link voltage variable that is not present in the other equations. It is therefore impossible to derive one equation from the rest, because at least one variable cannot be accounted for. This proves that the set of m KVL equations is indeed independent and completes the proof.

The above criteria can be simplified somewhat if the network is *planar,* i.e., if the network graph can be drawn on a plane without any branches intersecting except at nodes. A network which does not satisfy this property is said to be *nonplanar*. To illustrate the difference between a planar and a nonplanar network, consider first the bridge network shown earlier in Fig. 4-2. This network is clearly planar because its graph has no intersecting branches. The lattice network shown in Fig. 4-4*a* looks like a nonplanar network, but do not be fooled by what appear to be intersecting branches. This network is planar because its graph can be redrawn in planar form as shown in Fig. 4-4*b*. However, the network shown in Fig. 4-5 is nonplanar because it is impossible to draw its graph on a plane without intersecting branches.

If a network is planar, then its graph can be stretched (imagine that the branches are made of rubber bands) into something that resembles a fish net. This leads us to define a *mesh* of a planar network to be a closed loop which encloses no branches in its interior. Observe that the term mesh is defined only for planar



(a) (b)

Fig. 4-4. A lattice network is planar because its graph can be drawn on a plane without intersecting branches.

networks and that while every mesh is a loop, not every loop is a mesh. For example, consider the planar graph of Fig. 4-2*b*. The loop formed by branches 2, 3, 4 is a mesh, but the loop formed by branches 2, 3, 8 is not because it contains branch 4 in its interior. The significance for defining meshes is that it leads naturally to the following simple criteria.

INDEPENDENT KVL MESH EQUATION CRITERIA

A planar network with n nodes and b branches contains exactly $b - (n - 1)$ meshes. Moreover, the KVL equations written around these meshes are independent.

The first part of the criteria can be proved rigorously by applying the principles of mathematical induction similar to the way we did earlier in proving property 1. The second part of these criteria can be proved by the method of contradiction. Since the details are quite lengthy, although not difficult, we shall leave the proof for the problems.

Exercise 1: Find the $b - (n - 1)$ fundamental loops associated with each of the four trees shown in Fig. 4-3.

Exercise 2: (*a*) Obtain the set of independent KVL equations associated with the tree T_1 of Fig. 4-3. (*b*) Add a KVL equation (written around another loop) to the equations already obtained in (*a*). Show that these equations are not independent by following the steps given in the proof of the independent KVL equation criteria.

Exercise 3: (*a*) Verify that the networks shown in Figs. 4-2 and 4-4 have exactly $b - (n - 1)$ meshes. (*b*) Show that the choice of these meshes is not unique by redrawing the associated graphs into as many distinct configurations as possible.

4-3-2 CRITERIA FOR WRITING INDEPENDENT KCL EQUATIONS

Our objective here is to formulate an analogous set of criteria for writing a system of independent KCL equations. Before we do this, it is necessary to introduce the concept of a cut set.

Given a network graph G, a *cut set* is a set of branches of G having the property that if we cut each branch in the set once (as if with scissors), the graph becomes separated into two disconnected subgraphs, and if we leave any one branch of the set uncut, the graph remains connected in one piece by that branch.

For example, consider the graph shown in Fig. 4-2*b*. The set of branches $\{3,4,8,6\}$ is a cut set, since cutting these branches once separates the graph into two subgraphs. Similarly, the set of branches $\{3,4,8,5,7\}$ is also a cut set. However, the set of branches $\{3,4,6\}$ is not a cut set because the graph is still connected by branch 8. The set of branches $\{1,2,3,4\}$ is not a cut set, but for a different reason. Even though cutting all four branches results in two separate subgraphs, it is not a cut set because it includes one more branch than is necessary to separate two subgraphs. Only the three branches $\{1,2,3\}$ which would still separate the graph into two subgraphs is a cut set. From this, we can conclude that the set of branches connected to any node is automatically a cut set.

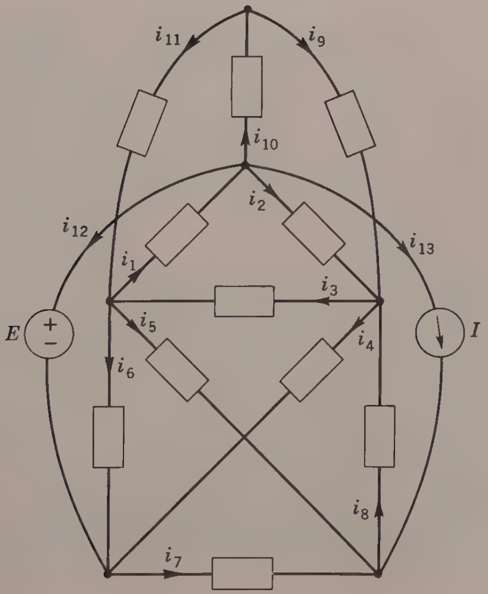


Fig. 4-5. An example of a non-planar network.

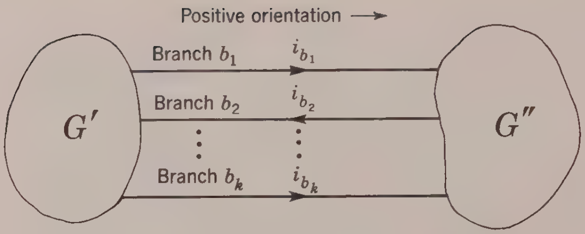


Fig. 4-6. The branches belonging to a cut set can always be redrawn to show the two subgraphs G' and G'' explicitly.

If we actually "cut" the branches belonging to a cut set, then these branches would be separated into two separate subgraphs. For convenience, we shall arbitrarily call them the *first subgraph* and the *second subgraph.* We shall next define a *positive orientation* for the cut set to be from the first subgraph to the second subgraph. If the current in any branch belonging to this cut set is in the same orientation, it is said to be positively directed. Otherwise, it is negatively directed. With these preliminaries we are now in a position to derive a more general form of Kirchhoff current law.

GENERALIZED KIRCHHOFF CURRENT LAW

The algebraic sum of all branch currents belonging to a cut set is zero. The sign of each current is assigned in accordance with the rule given above.

To prove this important result, let the branches $\{b_1,b_2, \ldots ,b_k\}$ be a cut set of an arbitrary network graph. In view of the definition of a cut set, we can always redraw the network graph in the form shown in Fig. 4-6 to indicate explicitly that the cutting of branches $\{b_1, \ldots ,b_k\}$ results in separating the network into two subgraphs G' and G''. Let us arbitrarily define the positive orientation of this cut set to be from G' to G''. To be specific, let us suppose that the subgraph G'' contains m nodes, where $m \geq 1$. When $m = 1$, the subgraph G'' would simply be a node connecting the branches $\{b_1, \ldots ,b_k\}$, and the theorem is automatically satisfied since KCL is then applicable at the node. If $m > 1$, then there will be more than one node inside G'' which connects not only the branches $\{b_1, \ldots ,b_k\}$, but also other internal branches. Suppose, corresponding to each internal node in G'', we write the KCL equation at that node.[1] This would give us m KCL equations. Observe now that each variable i_j corresponding to an internal branch of G'' would appear in exactly two of the m KCL equations and with

---

[1] We assume here that care has been taken to ensure that current entering each node is consistently assumed to be positive or negative. This is necessary for the "cancellation" required in the latter part of the proof to take place.

opposite signs. This must be so because each internal branch of G'' must be connected between two internal nodes x and y of G''. The current corresponding to this branch is directed either from node x to node y or vice versa, but in any case the KCL equations written at node x and at node y would each contain i_j with opposite signs. The variables that will appear only once in the above set of KCL equations will be the currents $i_{b_1}, i_{b_2}, \ldots, i_{b_k}$. Hence, if we add the corresponding sides of all the m KCL equations, the internal currents in the left side of the resulting equation would cancel each other out, leaving the equation

$$i_{b_1} - i_{b_2} + \cdots + i_{b_k} = 0$$

This is equivalent to saying that the algebraic sum of currents in all branches of any cut set is zero. Hence, the result is proved.

Observe that the generalized KCL includes the original version as a special case. Actually, it is sometimes useful to imagine each of the two subgraphs G' and G'' in Fig. 4-6 as a "supernode." In this case, the generalized KCL can be simplified as follows: The algebraic sum of all branch currents entering a supernode is zero.

We have seen earlier that a network can have many distinct cut sets. Although the generalized KCL is valid at each cut set, not all equations obtained this way are independent. The key to the criteria we are looking for lies in the recognition of a special class of cut sets known as the *fundamental cut sets.* A cut set corresponding to a tree T of a network graph is said to be a fundamental cut set if the branches belonging to the cut set contain only one tree branch of T and if the rest of the branches are links.

To illustrate the difference between a fundamental cut set and an ordinary cut set, consider again the network graph shown in Fig. 4-2*b*. If we pick the tree T_1 in Fig. 4-3 again, then the set of branches $\{1,7\}$ is a fundamental cut set. Similarly, the sets $\{7,2,3\}$, $\{7,5,8,4,3\}$, and $\{7,5,6\}$ are also fundamental cut sets, since each contains only one tree branch of T_1. Observe, however, that the set of branches $\{1,2,3\}$ is not a fundamental cut set corresponding to the tree T_1 because it contains two tree branches, 1 and 2.

In view of property 1 discussed earlier, it is clear that, corresponding to any tree, there are exactly $n - 1$ fundamental cut sets in any network graph. Or another way of saying it is that there are as many fundamental cut sets as there are branches in a tree.

We are now ready to state a fundamental result which shows us how to write a set of independent, generalized KCL equations.

INDEPENDENT KCL EQUATION CRITERIA

Let G be a network graph containing n nodes and b branches. Let T be any tree in G. Then *the maximum number of independent generalized KCL equations for G is $n - 1$. Moreover, these $n - 1$ independent equations can always be written by applying the generalized KCL to the $n - 1$ fundamental cut sets associated with the tree T.*

This important result can be proved in a manner very similar to the way the independent KVL equation criteria were proved. Hence, we shall leave the details of the proof for the problems.

Exercise 1: (*a*) List all the cut sets of the graph shown in Fig. 4-2*b*. (*b*) Redraw each of the cut sets obtained in (*a*) into the form of two subgraphs shown in Fig. 4-6. (*c*) Verify that each subgraph can be considered as a supernode for the purpose of applying KCL.

Exercise 2: (*a*) Find all fundamental cut sets of the network graph shown in Fig. 4-2*b* corresponding to each of the trees T_2, T_3, and T_4 listed in Fig. 4-3. (*b*) Verify that there are exactly $n - 1$ fundamental cut sets in each case.

4-4 EQUATIONS OF MOTION FOR RESISTIVE NETWORKS

The materials in the preceding section have been presented on rather formal grounds because of their generalities. To offer some relief from this overdose, we shall now be more specific. In particular, we shall consider several representative types of networks in their order of complexity. Our objective in this section is to go through these networks in great detail and show how to obtain an independent set of equations of motion for each network. In order not to distract our line of thought from this main objective, we shall go as far as simplifying the equations to a reasonably compact form, but we shall not solve them. The means for solving these equations will be discussed in Sec. 4-5.

4-4-1 NETWORKS CONTAINING ONLY TWO-TERMINAL RESISTORS AND INDEPENDENT SOURCES

The materials on network topology are clearly applicable to this class of networks. We shall consider two examples.

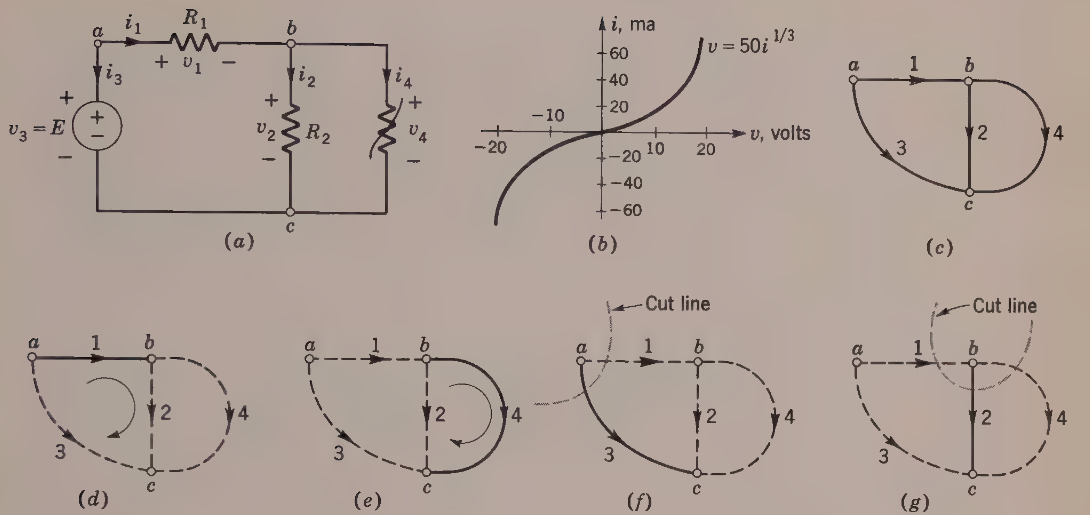


Fig. 4-7. A voltage stabilizer network and its associated fundamental loops and fundamental cut sets.

EXAMPLES

1. Consider the simple network shown in Fig. 4-7*a*. The only nonlinear element is a varistor with a v-i curve represented by $v = 50i^{1/3}$, where v is in volts and i is in amperes.

An inspection of this v-i curve in Fig. 4-7*b* reveals that for a certain range of currents the voltage varies only slightly. Because of this property, this circuit is usually used in practice with R_2 as the load resistor. Since the same voltage v appears across R_2, the load voltage is thereby stabilized against any undesirable fluctuation in the input source voltage E. As such, this circuit is usually called a *voltage-stabilizer network.* In order to design this circuit properly, it is necessary to solve for the voltage v_4 corresponding to each value of E. Our first step is to write the equations of motion describing this network.[1]

The graph of this network is shown in Fig. 4-7*c*. The equations of motion consist of two parts:[2]

Equations from the laws of elements

Element 1: $v_1 = R_1 i_1$ (4-6*a*)

Element 2: $v_2 = R_2 i_2$ (4-6*b*)

Element 3: $v_3 = E$ (4-6*c*)

Element 4: $v_4 = 50 i_4^{1/3}$ (4-6*d*)

[1] As we learn more techniques, we shall find many short cuts for solving this problem. However, our aim here is to illustrate the systematic methods for writing a system of independent equations. Therefore, we shall do it the hard way.

[2] Observe that it is important to choose the same unit for all current and voltage variables when writing the equations of motion.

Equations from the laws of interconnection

Since the graph has three nodes and four branches, $n = 3$, $b = 4$. Hence from the independent KVL equation criteria, we can write two independent equations by choosing two fundamental loops. Similarly, from the independent KCL equation criteria, we can write two independent KCL equations by choosing two fundamental cut sets. To do this, we must first pick a tree, say branches 2 and 3. The remaining branches 1 and 4 are therefore links. For our first fundamental loop, we pick link 1 and close the loop around tree branches 2 and 3, as shown in Fig. 4-7*d*. For the second fundamental loop, we pick link 4 and close the loop around tree branch 2, as shown in Fig. 4-7*e*. To obtain our first fundamental cut set, we pick tree branch 3 and form a cut set with link 1, as shown in Fig. 4-7*f*. To obtain the second fundamental cut set, we must pick tree branch 2 with links 1 and 4, as shown in Fig. 4-7*g*. The following independent equations from the laws of interconnection are now easily obtained:[1]

Fundamental loop	KVL equations	
Branches 1, 2, 3	$v_1 + v_2 - v_3 = 0$	(4-7*a*)
Branches 4, 2	$v_4 - v_2 = 0$	(4-7*b*)
Fundamental cut set	**KCL equations**	
Branches 3, 1	$i_3 + i_1 = 0$	(4-7*c*)
Branches 2, 1, 4	$i_2 - i_1 + i_4 = 0$	(4-7*d*)

Equations (4-6) and (4-7) together constitute a system of eight independent equations in terms of eight branch variables. In theory, these eight equations are the equations of motion of the network. In practice, however, it is always desirable that we try to eliminate as many variables as possible from these equations, so that the resulting equations will be simpler to solve. We can usually eliminate immediately at least half of the unknowns by substituting the equations from the laws of elements in terms of the respective variables into the laws of interconnection. Hence, substituting Eq. (4-6) into Eq. (4-7), we obtain the following four equations in four variables:

$$R_1 i_1 + R_2 i_2 = E \tag{4-8a}$$

$$R_2 i_2 = 50 i_4^{1/3} \tag{4-8b}$$

$$i_3 = -i_1 \tag{4-8c}$$

$$i_2 = (i_1 - i_4) \tag{4-8d}$$

---

[1] Unless specified otherwise, we shall pick the direction for which a fundamental loop is traced to be along the reference direction of the link in the loop. Similarly, the positive orientation of each cut set will be taken identical with the reference direction of the tree branch in the cut set.

We can usually reduce these equations further by an appropriate grouping of equations containing similar variables, thereby hoping that some of the variables can be solved in terms of the rest. From Eq. (4-8) we notice that the variable i_3 is present only in Eq. 4-8*c*. Hence, Eqs. (4-8*a*), (*b*), and (*d*) may be grouped together. A closer look at these equations suggests that i_2 can be eliminated if we substitute Eq. (4-8*d*) into Eqs. (4-8*a*) and (*b*); thus

$$R_1 i_1 + R_2(i_1 - i_4) = E \tag{4-9a}$$

$$R_2(i_1 - i_4) = 50 i_4^{1/3} \tag{4-9b}$$

These equations contain only two variables; consequently, it is even easier to see if any further elimination is possible. An inspection of these equations immediately suggests that we solve for i_1 in terms of i_4; thus

$$i_1 = \left(\frac{R_2}{R_1 + R_2}\right) i_4 + \left(\frac{E}{R_1 + R_2}\right) \tag{4-10}$$

Substituting Eq. (4-10) for i_1 in Eq. (4-9*b*) and rearranging the terms, we obtain, after some simplification, the relationship

$$i_4 = \frac{E}{R_1} - \left(\frac{R_1 + R_2}{R_1 R_2}\right)(50 i_4^{1/3}) \tag{4-11}$$

Equation (4-11) contains only one variable i_4. In some very special cases a further elimination may be possible, and the solution for i_4 found analytically. Unfortunately, it is safe to say that a complete analytical solution does not exist for most practical networks of interest. In fact, even for this simple network, Eq. (4-11) is just about as far as we can go. Other nonanalytical techniques must, therefore, be used to solve Eq. (4-11) for i_4. The solution will be presented in Sec. 4-5. Once i_4 is found, it can be substituted back into Eq. (4-10) to obtain the solution for i_1. Substituting the values of i_1 and i_4 into Eqs. (4-8*c*) and (4-8*d*), we obtain the solutions for i_3 and i_2. Finally, the respective voltages v_1, v_2, and v_4 can now be calculated from Eq. (4-6).

2. Consider next the more complicated network shown in Fig. 4-8*a*. This type of network is used in many important automatic controlled applications. The elements consist of a 50-kΩ linear resistor R_1, two identical monotonic resistors[1] R_2 and R_6 characterized by $v_2 = i_2^3$ and $v_6 = i_6^3$ (Fig. 4-8*b*); two identical neon bulbs R_3 and R_5 with current-controlled *v-i* curves $v_3 = f(i_3)$ and $v_5 = f(i_5)$ (Fig. 4-8*c*); a tunnel diode with a

---

[1] A resistor is said to be monotonic if it is characterized by a strictly monotonically increasing *v-i* curve.

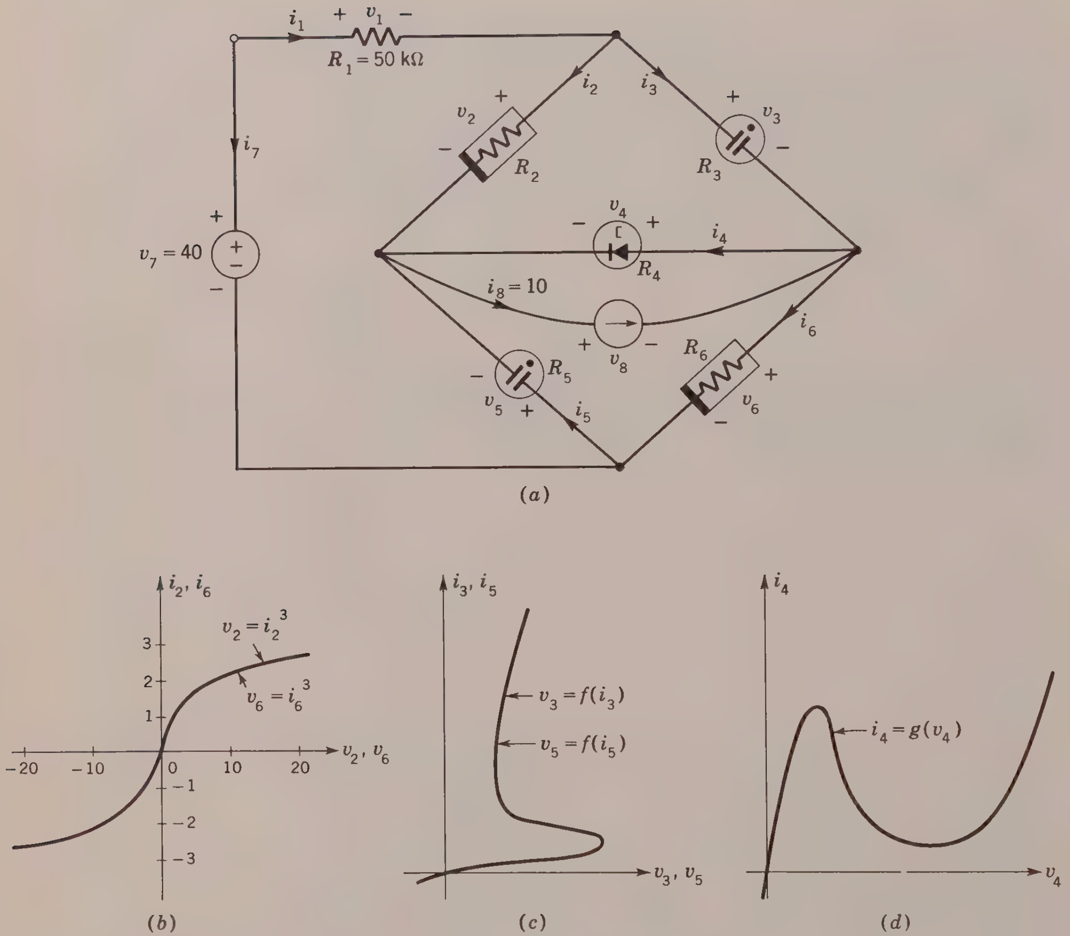


Fig. 4-8. A nonlinear bridge network containing both current-controlled and voltage-controlled resistors.

voltage-controlled v-i curve $i_4 = g(v_4)$ (Fig. 4-8d); and two independent sources.

The equations of motion, as usual, must consist of two parts.

Equations from the laws of elements

Element 1: $v_1 = 50i_1$ (4-12a)

Element 2: $v_2 = i_2^3$ (4-12b)

Element 3: $v_3 = f(i_3)$ (4-12c)

Element 4: $i_4 = g(v_4)$ (4-12d)

Element 5: $v_5 = f(i_5)$ (4-12*e*)

Element 6: $v_6 = i_6{}^3$ (4-12*f*)

Element 7: $v_7 = 40$ (4-12*g*)

Element 8: $i_8 = 10$ (4-12*h*)

Equations from the laws of interconnection

A comparison between the present network and that shown earlier in Fig. 4-2*a* shows that their network graphs are identical. Referring to this graph in Fig. 4-2*b*, we found $n = 5$, $b = 8$, $b - (n - 1) = 4$, and $(n - 1) = 4$. Hence, applying the independent KVL and KCL equation criteria, we know there are exactly four independent KVL equations and four independent KCL equations. To write these equations, we arbitrarily pick the tree T_1 consisting of branches 1, 2, 8, 6 (Fig. 4-3). The following equations corresponding to each fundamental loop and cut set (with respect to the tree T_1) can be easily obtained by inspection:

Fundamental loop	KVL equations	
Branches 7, 6, 8, 2, 1	$v_7 - v_6 - v_8 - v_2 - v_1 = 0$	(4-13*a*)
Branches 3, 8, 2	$v_3 - v_8 - v_2 = 0$	(4-13*b*)
Branches 4, 8	$v_4 + v_8 = 0$	(4-13*c*)
Branches 5, 8, 6	$v_5 + v_8 + v_6 = 0$	(4-13*d*)
Fundamental cut set	KCL equations	
Branches 1, 7	$i_1 + i_7 = 0$	(4-13*e*)
Branches 2, 3, 7	$i_2 + i_3 + i_7 = 0$	(4-13*f*)
Branches 8, 3, 4, 5, 7	$i_8 + i_3 - i_4 - i_5 + i_7 = 0$	(4-13*g*)
Branches 6, 5, 7	$i_6 - i_5 + i_7 = 0$	(4-13*h*)

Equations (4-12) and (4-13) constitute a system of 16 independent equations in terms of 16 branch variables. As usual, we can immediately eliminate half of these variables by substituting Eq. (4-12) into Eq. (4-13) and obtain the following eight independent equations in eight variables:

$$40 - i_6{}^3 - v_8 - i_2{}^3 - 50i_1 = 0 \qquad \text{(4-14}a\text{)}$$

$$f(i_3) - v_8 - i_2{}^3 = 0 \qquad \text{(4-14}b\text{)}$$

$$v_4 + v_8 = 0 \qquad \text{(4-14}c\text{)}$$

$$f(i_5) + v_8 + i_6{}^3 = 0 \qquad \text{(4-14}d\text{)}$$

$$i_1 + i_7 = 0 \qquad \text{(4-14}e\text{)}$$

$$i_2 + i_3 + i_7 = 0 \tag{4-14f}$$

$$10 + i_3 - g(v_4) - i_5 + i_7 = 0 \tag{4-14g}$$

$$i_6 - i_5 + i_7 = 0 \tag{4-14h}$$

These equations can be further reduced by observing that Eqs. (4-14*c*), (*e*), (*f*), and (*h*) are linear equations and can be solved for

$$v_8 = -v_4$$

$$i_7 = -i_1$$

$$i_2 = -i_3 - i_7 = -i_3 + i_1$$

$$i_6 = i_5 - i_7 = i_5 + i_1$$

Substituting these equations for the corresponding variables in Eq. (4-14), we obtain

$$40 - (i_1 + i_5)^3 + v_4 - (i_1 - i_3)^3 - 50i_1 = 0 \tag{4-15a}$$

$$f(i_3) + v_4 - (i_1 - i_3)^3 = 0 \tag{4-15b}$$

$$f(i_5) - v_4 + (i_1 + i_5)^3 = 0 \tag{4-15c}$$

$$10 + i_3 - g(v_4) - i_5 - i_1 = 0 \tag{4-15d}$$

We can simplify these equations further by solving for v_4 from Eq. (4-15*b*) and substituting it into Eqs. (4-15*a*), (*c*), and (*d*). After some simplification, we obtain the following three independent equations in the three variables i_1, i_3, and i_5:

$$40 - (i_1 + i_5)^3 - f(i_3) - 50i_1 = 0 \tag{4-16a}$$

$$f(i_5) - (i_1 - i_3)^3 + f(i_3) + (i_1 + i_5)^3 = 0 \tag{4-16b}$$

$$10 + i_3 - g[(i_1 - i_3)^3 - f(i_3)] - i_5 - i_1 = 0 \tag{4-16c}$$

The expression inside the brackets in Eq. (4-16*c*) is the argument of the function $g(v)$. We would like to eliminate some more variables if possible, but again we seem to have reached the end of the "analytic" road. We shall, therefore, stop here and postpone until Sec. 4-5 a discussion of other avenues for solving these equations.

Exercise 1: Obtain the equations of motion for the network shown in Fig. 4-7*a* by choosing a different tree consisting of branches 3 and 4. Simplify the equations to arrive at Eq. (4-11).

Exercise 2: Find the equations of motion of the network shown in Fig. 4-7*a* but with the load resistor R_2 replaced by a vacuum diode (represented by a three-half power law $i_2 = kv_2^{3/2}$). Simplify the equations as far as you can go.

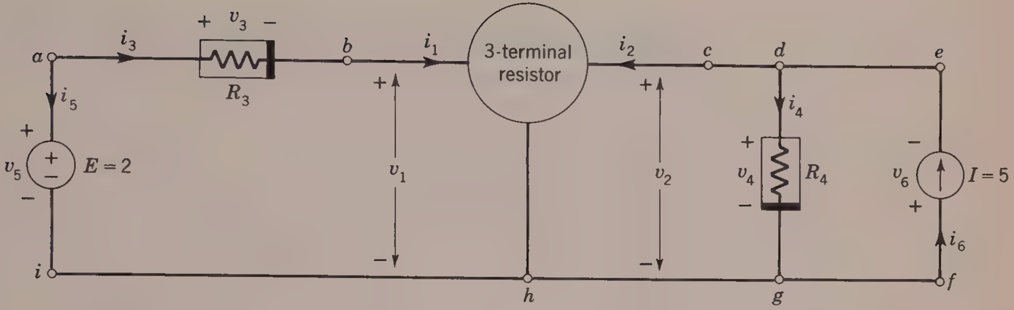


Fig. 4-9. A nonlinear network containing a hypothetical three-terminal resistor and two 2-terminal resistors.

Exercise 3: Derive Eq. (4-16) of the network shown in Fig. 4-8*a* by choosing a different tree consisting of branches 1, 3, 6, 5.

4-4-2 NETWORKS CONTAINING MULTITERMINAL RESISTORS AND CONTROLLED SOURCES

When a resistive network contains n-terminal resistors ($n > 2$) or controlled sources or both, most of the topological results presented in Sec. 4-3 are not applicable in their original forms. In many cases, these results can be extended under appropriate modifications. In general, however, these modifications are too complex to be of much practical value, and hence we shall not dwell upon this subject any longer.[1] Our techniques for proving the topological results in Sec. 4-3 are not completely useless, however, for by applying the same principles, it is usually possible to arrive at an independent system of equations for most practical networks of interest. Perhaps a couple of examples will illustrate the general procedure. In order not to get ourselves bogged down in the graphical details normally required with any practical device, we have chosen hypothetical devices for the following examples.

EXAMPLES

1. Consider the network shown in Fig. 4-9. The two resistors R_3 and R_4 are assumed to be characterized by $i_3 = 3 + v_3^3 - e^{-v_3}$ and $i_4 = 1 + v_4^3 - 2v_4^2 + 5v_4^7$, respectively. The three-terminal resistor is characterized by $i_1 = 1 + v_1 + 3v_1 i_2^3 - 4i_2^5$ and $v_2 = 4 - i_2 v_1 - 2i_2^2 v_1^5 + v_1^3$. The units of all voltages are in volts, and the units of all currents are in amperes.

   There is a total of 12 variables in this network, namely, $v_1, v_2, \ldots, v_6$ and $i_1, i_2, \ldots, i_6$. Hence, we must obtain 12 independent equations in order to solve for these variables. As usual, the equations consist of two parts:

---

[1] Simple, practical criteria analogous to the independent KVL and KCL equation criteria would have been very useful. Unfortunately, no such results are presently available, and it is doubtful that they can be found without greatly restricting the class of allowable networks.

Equations from the laws of elements

Three-terminal resistor:
$$\begin{cases} i_1 = 1 + v_1 + 3v_1 i_2^3 - 4i_2^5 & \text{(4-17}a\text{)} \\ v_2 = 4 - i_2 v_1 - 2i_2^2 v_1^5 + v_1^3 & \text{(4-17}b\text{)} \end{cases}$$

Resistor R_3: $$i_3 = 3 + v_3^3 - e^{-v_3} \qquad \text{(4-17}c\text{)}$$

Resistor R_4: $$i_4 = 1 + v_4^3 - 2v_4^2 + 5v_4^7 \qquad \text{(4-17}d\text{)}$$

Element 5: $$v_5 = 2 \qquad \text{(4-17}e\text{)}$$

Element 6: $$i_6 = 5 \qquad \text{(4-17}f\text{)}$$

We already know that the equations from the laws of elements are always independent. Therefore, any dependency must come from the laws of interconnection. A review of the proofs of the various results in Sec. 4-3 shows that as long as each KVL or KCL equation is written such that it contains at least one new variable which did not appear in the previously written KVL or KCL equations, then no dependency can occur. In particular, if the network is planar, then the KVL equations around each mesh would certainly be independent because each new mesh must certainly pick up at least a new branch. As for KCL equations, a simple procedure consists of choosing cut sets so that each cut set contains at least one branch not included in the previous cut sets. For this example, the network is planar, and hence we shall simply write KVL equations around the three meshes formed by nodes *a*, *b*, *h*, *i*; nodes *c*, *d*, *g*, *h*; and nodes *d*, *e*, *f*, *g*. Similarly, the branches terminating at nodes *a*, *b*, and *d* constitute three independent cut sets. The respective equations are as follows.

Equations from the laws of interconnection

Independent loop formed by	KVL equations	
Nodes *a*, *b*, *h*, *i*	$v_5 - v_1 - v_3 = 0$	(4-18*a*)
Nodes *c*, *d*, *g*, *h*	$v_2 - v_4 = 0$	(4-18*b*)
Nodes *d*, *e*, *f*, *g*	$v_4 + v_6 = 0$	(4-18*c*)
Independent cut set formed by	**KCL equations**	
Branches 3, 5	$i_3 + i_5 = 0$	(4-18*d*)
Branches 1, 3	$i_1 - i_3 = 0$	(4-18*e*)
Branches 2, 4, 6	$i_2 + i_4 - i_6 = 0$	(4-18*f*)

Equations (4-17) and (4-18) constitute 12 independent equations in 12 variables. As usual, we can immediately eliminate half of these variables by substituting Eq. (4-17) into Eq. (4-18); thus

$$2 - v_1 - v_3 = 0 \tag{4-19a}$$

$$(4 - i_2v_1 - 2i_2^2v_1^5 + v_1^3) - v_4 = 0 \tag{4-19b}$$

$$v_4 + v_6 = 0 \tag{4-19c}$$

$$(3 + v_3^3 - e^{-v_3}) + i_5 = 0 \tag{4-19d}$$

$$(1 + v_1 + 3v_1i_2^3 - 4i_2^5) - (3 + v_3^3 - e^{-v_3}) = 0 \tag{4-19e}$$

$$i_2 + (1 + v_4^3 - 2v_4^2 + 5v_4^7) - 5 = 0 \tag{4-19f}$$

These equations can be further simplified by observing that we can solve v_3, v_4, v_6, and i_5 in terms of the remaining variables; thus

$$\text{Eq. (4-19}a) \rightarrow v_3 = (2 - v_1) \tag{4-20a}$$

$$\text{Eq. (4-19}b) \rightarrow v_4 = (4 - i_2v_1 - 2i_2^2v_1^5 + v_1^3) \tag{4-20b}$$

$$\text{Eq. (4-19}c) \rightarrow v_6 = -v_4 \tag{4-20c}$$

$$\text{Eq. (4-19}d) \rightarrow i_5 = -(3 + v_3^3 - e^{-v_3}) \tag{4-20d}$$

Substituting Eqs. (4-20*a*) to (*d*) into Eqs. (4-19*e*) and (*f*), we obtain, after some simplification,[1]

$$v_1 + 3v_1i_2^3 - 4i_2^5 - (2 - v_1)^3 + e^{-(2-v_1)} - 2 = 0 \tag{4-21a}$$

$$\begin{aligned} i_2 + (4 - i_2v_1 - 2i_2^2v_1^5 + v_1^3)^3 - 2(4 - i_2v_1 - 2i_2^2v_1^5 + v_1^3)^2 \\ + 5(4 - i_2v_1 - 2i_2^2v_1^5 + v_1^3)^7 - 4 = 0 \end{aligned} \tag{4-21b}$$

These two equations cannot be further reduced and must be solved by other nonanalytic methods to be discussed in Sec. 4-5. Once the values of v_1 and i_2 are found, the values of v_3 and v_4 can be obtained from Eqs. (4-20*a*) and (*b*). Similarly, v_6 and i_5 can be found from Eqs. (4-20*c*) and (*d*). The remaining variables can be found from Eq. (4-17).

2. As a final example, let us consider a network with a controlled source, an ideal transformer, and a gyrator, in addition to two-terminal and three-terminal resistors, as shown in Fig. 4-10. For the sake of generality, we represent the three-terminal resistor by two functional equations $i_5 = g_1(v_5,v_6)$ and $i_6 = g_2(v_5,v_6)$. The two resistors R_2 and R_9 are represented by $i_2 = 5v_2^{1/3}$ and $v_9 = i_9 - 3i_9^5$, respectively. The terminal voltage of the voltage source is assumed to be equal to 10 sin $\omega_0 t$. As usual, the equations of motion consist of two parts.

---

[1] It is important to observe that the equations of motion for nonlinear networks containing nonmonotonic elements will generally contain both current and voltage branch variables.

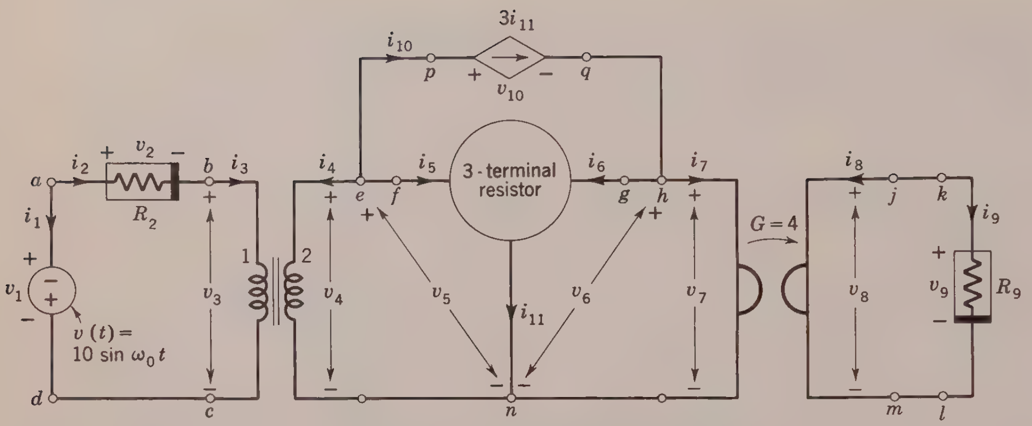


Fig. 4-10. A nonlinear network containing a controlled source, an ideal transformer, and a gyrator.

Equations from the laws of elements

Element 1.†	$v_1 = -10 \sin \omega_0 t$	(4-22*a*)
Resistor 2:	$i_2 = 5v_2^{1/3}$	(4-22*b*)
Ideal transformer:	$i_3 = -2i_4$	(4-22*c*)
	$v_4 = 2v_3$	(4-22*d*)
Three-terminal resistor:	$i_5 = g_1(v_5, v_6)$	(4-22*e*)
	$i_6 = g_2(v_5, v_6)$	(4-22*f*)
Gyrator:	$i_7 = 4v_8$	(4-22*g*)
	$i_8 = -4v_7$	(4-22*h*)
Resistor 9:	$v_9 = i_9 - 3i_9^5$	(4-22*i*)
Controlled source:	$i_{10} = 3i_{11}$	(4-22*j*)

Equations from the laws of interconnection

Independent loop formed by:	KVL equations	
Nodes a, b, c, d	$v_1 - v_2 - v_3 = 0$	(4-23*a*)
Nodes e, f, n	$v_4 - v_5 = 0$	(4-23*b*)
Nodes g, h, n	$v_6 - v_7 = 0$	(4-23*c*)
Nodes j, k, l, m	$v_8 - v_9 = 0$	(4-23*d*)
Nodes e, p, q, h, n	$v_{10} + v_6 - v_5 = 0$	(4-23*e*)
Independent cut set formed by:	**KCL equations**	
Branches 1, 2	$i_1 + i_2 = 0$	(4-23*f*)
Branches 2, 3	$i_2 - i_3 = 0$	(4-23*g*)
Branches 4, 5, 10	$i_4 + i_5 + i_{10} = 0$	(4-23*h*)
Branches 6, 7, 10	$i_6 + i_7 - i_{10} = 0$	(4-23*i*)
Branches 8, 9	$i_8 + i_9 = 0$	(4-23*j*)
Branches 11, 4, 7	$i_{11} + i_4 + i_7 = 0$	(4-23*k*)

† To demonstrate that the reference polarity or direction can be arbitrarily assigned, we define v_1 to have a polarity opposite to the voltage source in Fig. 4-10. Accordingly, the law of this element must be multiplied by -1.

It is easy to see that Eqs. (4-23*a*) to (*e*) and Eqs. (4-23*f*) to (*k*) are independent because each equation contains at least one variable which did not appear earlier. Therefore, Eqs. (4-22) and (4-23) together constitute a system of 21 independent equations in 21 variables. This odd number of equations arises because the current-controlled current source is controlled by a different variable i_{11} which did not previously appear in the element variables. Corresponding to the 10 equations in Eq. (4-22), we can eliminate 10 variables by substituting these equations into Eq. (4-23) to obtain

$$-10 \sin \omega_0 t - v_2 - v_3 = 0 \tag{4-24a}$$

$$2v_3 - v_5 = 0 \tag{4-24b}$$

$$v_6 - v_7 = 0 \tag{4-24c}$$

$$v_8 - (i_9 - 3i_9{}^5) = 0 \tag{4-24d}$$

$$v_{10} + v_6 - v_5 = 0 \tag{4-24e}$$

$$i_1 + (5v_2{}^{1/3}) = 0 \tag{4-24f}$$

$$5v_2{}^{1/3} - (-2i_4) = 0 \tag{4-24g}$$

$$i_4 + g_1(v_5,v_6) + (3i_{11}) = 0 \tag{4-24h}$$

$$g_2(v_5,v_6) + (4v_8) - (3i_{11}) = 0 \tag{4-24i}$$

$$(-4v_7) + i_9 = 0 \tag{4-24j}$$

$$i_{11} + i_4 + 4v_8 = 0 \tag{4-24k}$$

These equations can be further simplified by solving for a common set of variables in terms of those that are hard to solve. Hence, we obtain from Eq. (4-24) the following relationships:

$$\text{Eq. (4-24}b) \rightarrow v_3 = \tfrac{1}{2}v_5 \tag{4-25a}$$

$$\text{Eq. (4-24}c) \rightarrow v_7 = v_6 \tag{4-25b}$$

$$\text{Eq. (4-24}e) \rightarrow v_{10} = v_5 - v_6 \tag{4-25c}$$

$$\text{Eq. (4-24}f) \rightarrow i_1 = -5v_2{}^{1/3} \tag{4-25d}$$

$$\text{Eq. (4-24}g) \rightarrow i_4 = -\tfrac{5}{2}v_2{}^{1/3} \tag{4-25e}$$

$$\text{Eq. (4-24}j) \rightarrow i_9 = 4v_7 = 4v_6 \tag{4-25f}$$

$$\text{Eq. (4-24}d) \rightarrow v_8 = i_9 - 3i_9{}^5 = 4v_6 - 3(4v_6)^5 \tag{4-25g}$$

$$\text{Eq. (4-24}k) \rightarrow i_{11} = -i_4 - 4v_8 = \tfrac{5}{2}v_2{}^{1/3} - 4[4v_6 - 3(4v_6)^5] \tag{4-25h}$$

Observe that only three variables appear on the right sides of Eqs. (4-25*a*) to (*h*), namely, v_2, v_5, and v_6. Substituting these variables into the remaining equations in Eq. (4-24), we obtain

$$\text{Eq. (4-24}a) \rightarrow v_2 = -10 \sin \omega_0 t - \tfrac{1}{2} v_5 \tag{4-26a}$$

$$\text{Eq. (4-24}h) \rightarrow -\tfrac{5}{2} v_2^{1/3} + g_1(v_5,v_6) + \tfrac{15}{2} v_2^{1/3} - 12[4v_6 - 3(4v_6)^5] = 0 \tag{4-26b}$$

$$\text{Eq. (4-24}i) \rightarrow g_2(v_5,v_6) + 4[4v_6 - 3(4v_6)^5] - \tfrac{15}{2} v_2^{1/3} + 12[4v_6 - 3(4v_6)^5] = 0 \tag{4-26c}$$

These equations can be simplified further by substituting Eq. (4-26*a*) for v_2 in Eqs. (4-26*b*) and (*c*), and rearranging the terms; thus

$$5[-10 \sin \omega_0 t - \tfrac{1}{2} v_5]^{1/3} + g_1(v_5,v_6) - 12[4v_6 - 3(4v_6)^5] = 0 \tag{4-27a}$$

$$g_2(v_5,v_6) + 16[4v_6 - 3(4v_6)^5] - \tfrac{15}{2}(-10 \sin \omega_0 t - \tfrac{1}{2} v_5)^{1/3} = 0 \tag{4-27b}$$

This is as far as we can go. The solutions to these equations, of course, depend on the functional equations $g_1(v_5,v_6)$ and $g_2(v_5,v_6)$ of the three-terminal resistor. In general, Eqs. (4-27*a*) and (*b*) are highly nonlinear, and can only be found by the nonanalytic methods to be discussed in the next section. However, we realize that the readers must be itching by now for a complete solution in order to obtain a feeling of what it looks like. It seems appropriate, therefore, that we contrive a special case so that we may continue along the analytic road to home. Hence, let the three-terminal resistor be characterized by

$$i_5 = g_1(v_5,v_6) = 5(-\tfrac{1}{2} v_5)^{1/3} + 12[4v_6 - 3(4v_6)^5] \tag{4-28a}$$

$$i_6 = g_2(v_5,v_6) = -\tfrac{15}{2}[2v_6 + \tfrac{1}{2} v_5]^{1/3} - 16[4v_6 - 3(4v_6)^5] \tag{4-28b}$$

Substituting Eq. (4-28*a*) for $g_1(v_5,v_6)$ in Eq. (4-27*a*), we obtain, upon some simplification, the solution

$$v_5(t) = -10 \sin \omega_0 t \tag{4-29a}$$

Similarly, substituting Eq. (4-28*b*) for $g_2(v_5,v_6)$ in Eq. (4-27*b*), we obtain

$$v_6(t) = 5 \sin \omega_0 t \tag{4-29b}$$

Substituting Eq. (4-29*a*) for v_5 in Eq. (4-26*a*), we obtain

$$v_2(t) = -5 \sin \omega_0 t \tag{4-29c}$$

Upon substituting the solutions $v_2(t)$, $v_5(t)$, and $v_6(t)$ into Eqs. (4-25) and (4-22), the solutions for the remaining variables are trivially obtained. It is important to observe that the solutions of this network are all functions of time because of the presence of the ac voltage source.

Exercise 1: Derive Eq. (4-21) of the network shown in Fig. 4-9 by using the following loops and cut sets: (*a*) loops formed by nodes *a*, *b*, *h*, *i*; nodes *c*, *d*, *e*, *f*, *g*, *h*; and nodes *d*, *e*, *f*, *g*; (*b*) cut sets formed by branches 2, 4, 6; branches 5, 4, 6, and the common terminal of the three-terminal resistor; branches 1, 2, and the common terminal of the three-terminal resistor.

Exercise 2: (*a*) Find the solutions of the remaining variables of the network shown in Fig. 4-10 in terms of the solutions given by Eqs. (4-29*a*), (*b*), and (*c*). (*b*) Under what situations are the solutions of a resistive network independent of time?

4-5 PRACTICAL METHODS FOR SOLVING NONLINEAR FUNCTIONAL EQUATIONS

The equations of motion of any resistive nonlinear network always assume the form of a system of nonlinear functional equations. By applying standard elimination and substitution techniques, these equations can always be reduced to a smaller number of equations in terms of an equal number of variables. However, the examples in the preceding section clearly demonstrated the fact that there comes a point where no further simplification by these analytical techniques is possible.

Three nonanalytical methods are available for solving these equations, namely, *numerical*, *graphical*, and *piecewise-linear*. A common feature of these methods is that although conceptually quite simple, they are rather tedious to implement manually because a large number of repetitive calculations or graphical operations is generally required. However, because of their iterative nature, these methods are ideally suited for a digital computer. In fact, any calculation or operation which can be described by a step-by-step procedure (in the order of its occurrence) can be carried out by a computer. This step-by-step procedure is analogous to a "recipe" and is called an *algorithm*. Since this book does not assume any programming experience on the part of the readers, only algorithms associated with the various methods of analysis are emphasized. Accordingly, most of the examples and problems presented in this book were chosen to be simple enough so that they can be solved without a computer. However, the reader is strongly urged to learn how to write programs for implementing a given algorithm. Not only will he find this background to be invaluable, he will also be in a position to solve more realistic and complex problems.

An in-depth treatment of numerical methods would require some mathematical backgrounds beyond those assumed in this book. On the other hand, the remaining two methods, graphical and piecewise-linear, can be easily understood without any mathematical background beyond calculus. These two methods also give more insight into the actual operation of the circuits. Accordingly, a considerable portion of this book is devoted to an in-depth treatment of these methods.[1] In order to appreciate the power of numerical techniques, however, let us now turn to a study of some of the more basic concepts in numerical analysis.

4-5-1 THE FIXED-POINT CONCEPT

Perhaps the best way to introduce the powerful concept of a *fixed point* is through an example. Consider first the simplest case of one equation in one unknown, such as Eq. (4-11). Since any numerical method involves the operation of actual numbers, not symbols, let us arbitrarily choose $R_1 = R_2 = 50$ kΩ and $E = 200$ volts in Eq. (4-11) and obtain

$$i_4 = 4 - 2i_4^{1/3} \tag{4-30}$$

where i_4 is the current in milliamperes. To avoid writing down the subscript, let us choose the symbol X in place of i_4 and rewrite Eq. (4-30) as

$$X = F(X) \tag{4-31}$$

where

$$F(X) = 4 - 2X^{1/3} \tag{4-32}$$

Our object is to find a number $X = X^*$ which reduces Eq. (4-31) to an identity. Suppose we start by making an arbitrary initial guess $X = X_0$. In most likelihood, our guess is apt to be wrong in the sense that if we let $F(X_0) \equiv X_1$, then $X_0 \neq X_1$. If we were to make other guesses arbitrarily, chances are that we would never hit at the right value, unless we can find an algorithm which guarantees that each successive guess would be closer to the correct value of X^*.

It turns out that under appropriate conditions, the value $X_1 \equiv F(X_0)$ is a better guess than X_0 in the sense that the error is smaller; namely, $|X^* - X_1| < |X^* - X_0|$. This observation immediately suggests that if we make X_1 our next guess, then $X_2 \equiv F(X_1)$ would be an even better guess than X_1. If we keep repeating this procedure, we would eventually arrive at the correct value of X^*.

---

[1] The graphical method is discussed in detail in Chap. 6, and the piecewise-linear method in Chap. 12.

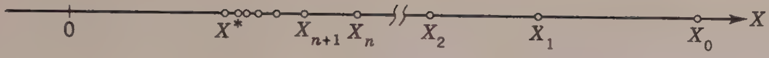


Fig. 4-11. If we interpret that each point X_n is being moved into the next point X_{n+1}, then the solution X^* corresponds to the point which is being moved into itself; namely, it is a fixed point.

Our algorithm consists, therefore, of the following steps:

Make an arbitrary initial guess $X = X_0$

Calculate $F(X_0)$ and call it X_1

Calculate $F(X_1)$ and call it X_2

Calculate $F(X_2)$ and call it X_3

. . . .

Calculate $F(X_n)$ and call it X_{n+1}

Stop whenever the difference between X_{n+1} and X_n becomes insignificant

This algorithm can be described compactly by the iteration formula

$$X_{n+1} = F(X_n) \qquad n = 0, 1, \ldots, n \tag{4-33}$$

If we interpret the value of X_n to be a point on the X axis, as shown in Fig. 4-11, then the algorithm above is equivalent to moving a point X_n into the next point X_{n+1}. If the distance between each successive points decreases rapidly enough, then it is clear that eventually we would converge at a point X^* such that each subsequent movement will be onto itself. In view of this interpretation, the point X^* is called a *fixed point* of the function $F(X)$. Accordingly, the above algorithm is called the *fixed-point algorithm.* This algorithm is of such fundamental importance that we shall summarize it as follows.

FIXED-POINT ALGORITHM

The solution of any equation written in the form $X = F(X)$ is equal to the fixed point of the function $F(X)$. Under appropriate conditions, the algorithm defined by Eq. (4-33) converges to the desired fixed point.[1]

Let us now apply the fixed-point algorithm to find the solution of Eq. (4-30); i.e., let us find the fixed point of Eq. (4-32). It should be clear that this task is ideally suited for a computer.[2] The results of carrying out this algorithm, either by hand or by a computer, corresponding to three different initial guesses are shown in

---

[1] It can be proved that a sufficient, but not necessary, condition to guarantee convergence is that the magnitude of the slope of $F(X)$ be less than unity; i.e., $|F'(X)| < 1$, for all values of X.

[2] Anyone with a little programming experience should be able to write a program to carry out the fixed-point iteration. For those who do not know programming, this is an excellent example and an excellent time to start.

TABLE 4-1 *The computer-printed output of the results of the fixed-point iteration of the equation* $F(X) = 4 - 2X^{1/3}$.

```
      X             F(X)             X             F(X)             X             F(X)
--------------------------    --------------------------    --------------------------
 27.0000        -2.0000         10.0000        -0.3089         -8.0000         8.0000
 -2.0000         6.5198         -0.3089         5.3519          8.0000         0.0000
  6.5198         0.2637          5.3519         0.5016          0.0000         3.9938
--------------------------    --------------------------    --------------------------
  0.2637         2.7175          0.5016         2.4109          3.9938         0.8268
  2.7175         1.2090          2.4109         1.3182          0.8268         2.1228
  1.2090         1.8694          1.3182         1.8071          2.1228         1.4296
--------------------------    --------------------------    --------------------------
  1.8694         1.5363          1.8071         1.5639          1.4296         1.7470
  1.5363         1.6923          1.5639         1.6785          1.7470         1.5913
  1.6923         1.6167          1.6785         1.6231          1.5913         1.6651
--------------------------    --------------------------    --------------------------
  1.6167         1.6527          1.6231         1.6496          1.6651         1.6295
  1.6527         1.6354          1.6496         1.6369          1.6295         1.6465
  1.6354         1.6437          1.6369         1.6430          1.6465         1.6383
--------------------------    --------------------------    --------------------------
  1.6437         1.6397          1.6430         1.6400          1.6383         1.6422
  1.6397         1.6416          1.6400         1.6414          1.6422         1.6404
  1.6416         1.6407          1.6414         1.6408          1.6404         1.6413
--------------------------    --------------------------    --------------------------
  1.6407         1.6411          1.6408         1.6411          1.6413         1.6408
  1.6411         1.6409          1.6411         1.6409          1.6408         1.6410
  1.6409         1.6410          1.6409         1.6410          1.6410         1.6409
--------------------------    --------------------------    --------------------------
  1.6410         1.6410          1.6410         1.6410          1.6409         1.6410
  1.6410         1.6410          1.6410         1.6410          1.6410         1.6410
```

Table 4-1. A total of 20 iterations are listed, corresponding to the initial guess $X_0 = 27$ (columns 1 and 2), $X_0 = 10$ (columns 3 and 4), and $X_0 = -8$ (columns 5 and 6), respectively. Notice that in each case, the iteration converges to the value $X = 1.6410$. Hence, the fixed point of $F(X)$ is equal to 1.6410. It is instructive to compare the values of X_n in Table 4-1 corresponding to each initial guess. Observe that a closer initial guess would require a fewer number of iterations.

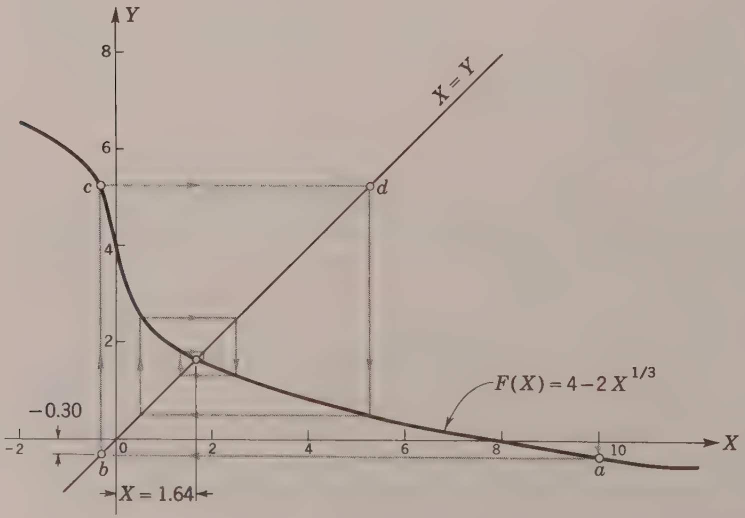
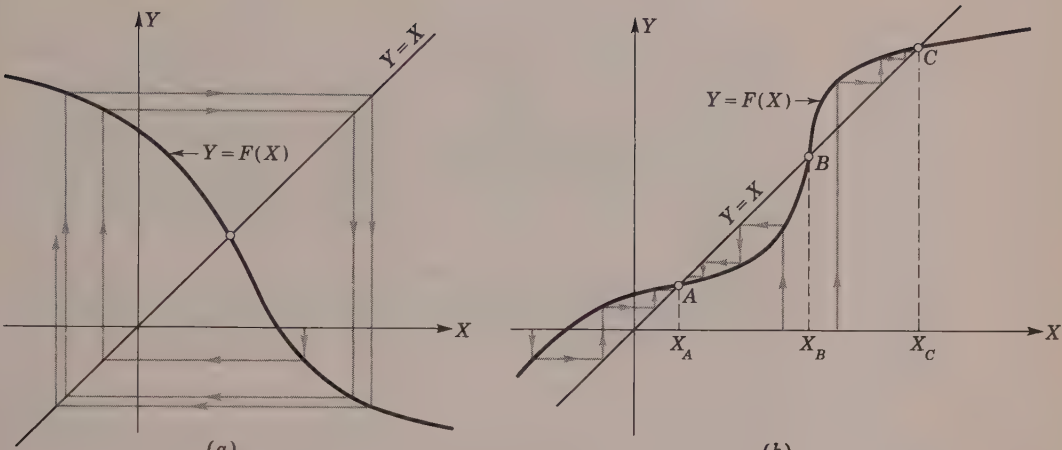


Fig. 4-12. A geometrical interpretation of the fixed-point algorithm.



(a)

(b)

Fig. 4-13. Depending on the nature of the curve $F(X)$, the fixed-point iteration (*a*) may not converge or (*b*) it may converge to only one of several possible fixed points.

The fixed-point algorithm has a simple geometrical interpretation. If we plot the curve $Y = F(X)$ as shown in Fig. 4-12 (corresponding to our present example), then the fixed point of $F(X)$ is simply the intersection between this curve and the 45° straight line through the origin. Since the ordinate (Y) and abscissa (X) of each point on this line are equal, the fixed-point algorithm is equivalent to the geometrical constructions shown in Fig. 4-12 (shown for the case $X_0 = 10$). Hence, through $X = 10$, a vertical line is dropped until it intersects the curve $F(X)$ at point *a*. If we project a horizontal line through *a* until it intersects the 45° line at point *b*, the abscissa ($X \approx -0.30$) at *b* gives the next value X_1. We can repeat the procedure by drawing a vertical line through *b* until it intersects $F(X)$ at point *c*. Projecting a horizontal line through *c* until it intersects the 45° line at point *d*, we obtain the value of X_2. Hence, this construction technique can be repeated a sufficient number of times until the points of intersection converge toward the fixed point.

The reader should verify that for this example, the above geometrical construction always converges to the desired fixed point, regardless of where we start. Unfortunately, we cannot guarantee that this behavior is true for all cases. As a matter of fact, there are two problems that one may encounter. The first problem is best seen from Fig. 4-13*a*. Observe that no matter where we start, the iteration diverges. The second problem occurs whenever the curve $F(X)$ has more than one point of intersection with the 45° line, as shown in Fig. 4-13*b*. In this case, it is clear that if the value of our initial guess X_0 is less than X_B, the iteration will converge to point A. Otherwise, it will converge to point C. Since it is not possible to converge to point B, this fixed point is said to be unstable. This example demonstrates a fundamental limitation of all numerical techniques; namely, the procedures cannot account for all solutions. Fortunately, we have two other methods which are independent of the number of solutions, namely, the graphical method (Chap. 6) and the piecewise-linear method (Chap. 12).†

To apply the fixed-point iteration technique, the equation must be of the form $X = F(X)$. In most problems, the equation is usually of the form

$$G(X) = 0 \tag{4-34}$$

It is always possible, however, to recast Eq. (4-34) into the proper form, namely,

$$X = X - G(X) \tag{4-35}$$

In this case, $F(X) = X - G(X)$. If the fixed-point iteration for $F(X)$ converges to a value X^*, then $X = X^*$ must be a solution of Eq. (4-34). If the iteration diverges, we can often correct the situation by multiplying $G(X)$ in Eq. (4-35) by an appropriate function $K(X)$, namely,

$$X = X - K(X)G(X) \tag{4-36}$$

Observe that if X^* is a solution of Eq. (4-34), then $G(X^*) = 0$ and Eq. (4-36) reduces to an identity. Hence, any solution of Eq. (4-34) is also a solution of Eq. (4-36). Conversely, suppose X^* is a fixed point of the function

$$F(X) = X - K(X)G(X) \tag{4-37}$$

then $X^* = X^* - K(X^*)G(X^*)$, or

$$K(X^*)G(X^*) = 0$$

If $K(X^*) \neq 0$, then we must have $G(X^*) = 0$. Hence, any fixed point X^* of Eq. (4-37) for which $K(X^*) \neq 0$ is also a solution of Eq. (4-34). This observation will allow us to choose suitable functions for $K(X)$ so that the fixed-point iteration converges. One such function is given in the next section.

---

† Resistive networks with multivalued solutions are commonly encountered in the process of analyzing switching networks. This subject will be discussed in detail in Chaps. 14 and 15.

Exercise 1: The equations of motion of a certain gamma-ray detection circuit have been reduced to

$$v = 2 + \left(\frac{v}{4}\right)^2$$

(*a*) Use the fixed-point algorithm to find the solution corresponding to the initial guesses $v = -4, 8, 12$, and 16. (*b*) Verify the solutions obtained in (*a*) graphically. (*c*) Draw the construction lines corresponding to each fixed-point iteration step in (*a*). Do this for all four initial guesses.

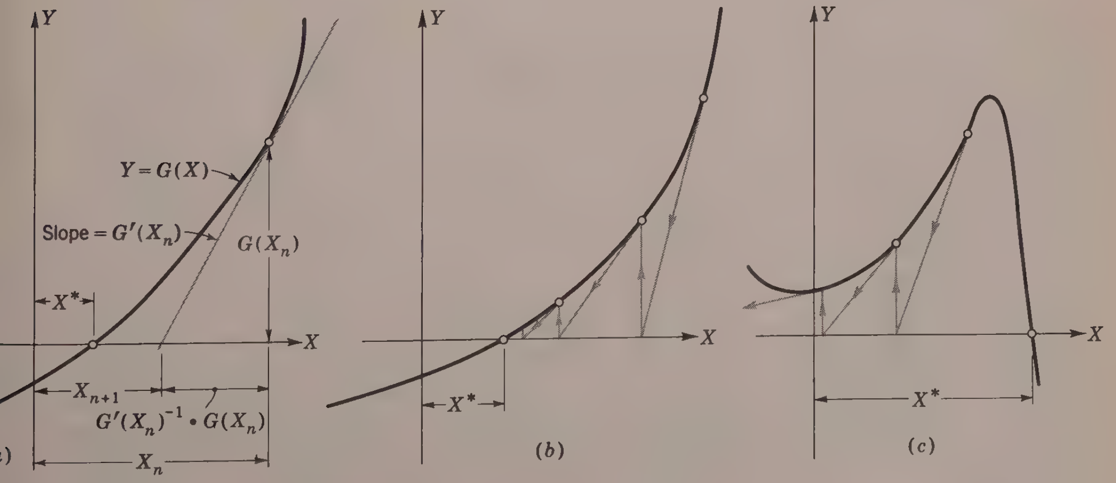
Exercise 2: In order to study the effect of the function $K(X)$ in Eq. (4-37) on the convergence of the fixed-point iteration, consider the equation $G(X) = 2X - 4 = 0$. (*a*) With $K(X) = 1$, perform the fixed-point iteration corresponding to the initial guess $X_0 = 6$. Does the iteration converge? (*b*) Repeat (*a*) with $K(X) = 1/4$.

Exercise 3: (*a*) Demonstrate geometrically that if the magnitude of the slope of $F(X)$ is less than unity for all values of X, then the fixed-point iteration is guaranteed to converge, regardless of the initial guess. (*b*) Show by a counterexample that the condition in (*a*) is not a necessary condition.

4-5-2 THE NEWTON-RAPHSON METHOD

We have demonstrated that the solutions of the nonlinear equation $G(X) = 0$ can be obtained by finding the fixed points of the associated function $F(X) = X - K(X)G(X)$. The function $K(X)$ (which does not vanish at the fixed points) is introduced to improve the rate of convergence. Among the many possible functions, the most common choice for $K(X)$ is the reciprocal of the derivative of $G(X)$, namely, $K(X) = [G'(X)]^{-1}$. In this case, Eqs. (4-37) and (4-33) become

Fig. 4-14. Geometrical interpretation of the Newton-Raphson method and its convergence behavior.



$$F(X) = X - [G'(X)]^{-1}G(X) \tag{4-38}$$

and

$$X_{n+1} = X_n - [G'(X_n)]^{-1}G(X_n) \tag{4-39}$$

The iteration method based on this fixed-point iteration formula is called the *Newton-Raphson method.* Equation (4-39) can be given a simple geometrical interpretation as shown in Fig. 4-14*a*. Corresponding to each value $X = X_n$, we draw a vertical line until it intersects the curve $Y = G(X)$. From this point of intersection, we draw a tangent line until it intersects the X axis. The value of X at this point gives the next value X_{n+1}. It is important to realize that the curve in Fig. 4-14*a* is $G(X)$ and not $F(X)$. Accordingly, the geometrical construction of the Newton-Raphson method is different from the fixed-point constructions presented earlier. The rate of convergence of the Newton-Raphson method also depends on the nature of the curve $G(X)$. In Fig. 4-14*b*, the convergence is very rapid; however, in Fig. 4-14*c* the iteration diverges. Observe that if the initial guess is near the correct solution X^*, then the Newton-Raphson method always converges. Therefore, if the iteration corresponding to a certain initial guess diverges, we can always choose some other guess until it converges. This is one property that makes the Newton-Raphson method a favorite choice.[1]

Let us apply the Newton-Raphson method to solve Eq. (4-30). This equation can be recast into the standard form

$$G(X) = X + 2X^{1/3} - 4 \tag{4-40}$$

The derivative of $G(X)$ is given by

$$G'(X) = 1 + \tfrac{2}{3}X^{-2/3}$$

Equation (4-39) can now be written as

$$X_{n+1} = F(X_n) \equiv X_n - \left(\frac{X_n + 2X_n{}^{1/3} - 4}{1 + \tfrac{2}{3}X_n{}^{-2/3}}\right) \tag{4-41}$$

For comparison purposes, we chose the same initial guesses as before, namely, $X_0 = 27$, 10, and -8. The results corresponding to 20 iterations are given in Table 4-2.[2] A comparison between the corresponding entries in Tables 4-1 and 4-2 shows that the Newton-

---

[1] Furthermore, it can be shown that the Newton-Raphson method converges faster than any other method of iteration.

[2] Since we did not know how fast the iteration would converge, we chose the same number of iterations (20) as before. Actually, it is quite easy to write a computer program to perform the above iterations under the additional instruction that the iteration be terminated as soon as the difference between two iterations reduces below a certain prescribed value.

Raphson method converges much faster than the earlier iteration method [with $K(X) = 1$].

Exercise 1: (*a*) Repeat Exercise 1 of Sec. 4-5-1 by the Newton-Raphson method. (*b*) Compare the rates of convergence of the two methods.

Exercise 2: (*a*) Repeat Exercise 2 of Sec. 4-5-1 by the Newton-Raphson method. (*b*) Compare the rates of convergence of the two methods.

4-5-3 EQUATIONS WITH MANY UNKNOWNS

The beauty of the fixed-point method lies in the fact that it can be easily generalized to solve a system of *n* nonlinear equations in *n* unknowns. To economize on notation, it suffices to consider the case of two equations in two unknowns; thus

$$\begin{aligned} X &= F_1(X,Y) \\ Y &= F_2(X,Y) \end{aligned} \tag{4-42}$$

The fixed point of Eq. (4-42) is any point (X^*, Y^*) which reduces the equations to an identity. By an analogous procedure, we can define the following iteration formula

$$\begin{aligned} X_{n+1} &= F_1(X_n, Y_n) \\ Y_{n+1} &= F_2(X_n, Y_n) \end{aligned} \tag{4-43}$$

TABLE 4-2 *The computer-printed output of the results of the Newton-Raphson iteration for $G(X) = X + 2X^{1/3} - 4$.*

```
      X            F(X)              X            F(X)              X            F(X)
 27.0000         0.0000                                          -8.0000        11.2000
  0.0000         0.0004          10.0000        0.9858           11.2000         0.8973
  0.0004         0.0297           0.9858        1.5977            0.8973         1.5810
  0.0297         0.4511           1.5977        1.6409            1.5810         1.6407
  0.4511         1.3956           1.6409        1.6410            1.6407         1.6410
  1.3956         1.6364           1.6410        1.6410            1.6410         1.6410
  1.6364         1.6410           1.6410        1.6410            1.6410         1.6410
  1.6410         1.6410           1.6410        1.6410            1.6410         1.6410
  1.6410         1.6410           1.6410        1.6410            1.6410         1.6410
  1.6410         1.6410           1.6410        1.6410            1.6410         1.6410
  1.6410         1.6410           1.6410        1.6410            1.6410         1.6410
  1.6410         1.6410           1.6410        1.6410            1.6410         1.6410
  1.6410         1.6410           1.6410        1.6410            1.6410         1.6410
  1.6410         1.6410           1.6410        1.6410            1.6410         1.6410
  1.6410         1.6410           1.6410        1.6410            1.6410         1.6410
  1.6410         1.6410           1.6410        1.6410            1.6410         1.6410
  1.6410         1.6410           1.6410        1.6410            1.6410         1.6410
  1.6410         1.6410           1.6410        1.6410            1.6410         1.6410
  1.6410         1.6410           1.6410        1.6410            1.6410         1.6410
  1.6410         1.6410           1.6410        1.6410            1.6410         1.6410
                                  1.6410        1.6410
```

Under appropriate conditions, the above iteration will converge to a fixed point of Eq. (4-42). In a similar manner, we can extend the Newton-Raphson method to the case of many unknowns. Since it involves some background in matrix analysis, we shall leave the details of this method to a more advanced course in numerical techniques.

Exercise 1: Consider the following system of two equations:

$X = Y^{1/3}$

$Y = X^{1/3}$

(*a*) Starting with the initial guess $(X_0, Y_0) = (8,27)$, apply Eq. (4-43) to find the fixed points of these equations. (*b*) Repeat (*a*) with $(X_0, Y_0) = (-27, -8)$. (*c*) Find the solutions analytically and discuss the general behavior of the above iteration method.

Exercise 2: Specify the iteration formula for finding the solution of a system of n equations with n unknowns $X_a, X_b, \ldots, X_n$.

Exercise 3: Transform the equations

$G_1(X,Y) = 0$

$G_2(X,Y) = 0$

into an equivalent fixed-point form similar to Eq. (4-37). Generalize this to a system of n equations in n unknowns.

4-6 EQUATIONS OF MOTION FOR DYNAMIC NETWORKS

By definition, a dynamic network must contain at least one capacitor or one inductor. Hence we expect the equations of motion of any dynamic network to contain at least one differential equation. Other than this difference, the procedure for writing the equations of motion of dynamic networks is similar to that of resistive networks. In fact, the equations obtained from the laws of interconnection would be identical to the equations obtained for a resistive network having an identical network graph. They will be a set of linear algebraic equations. This means that any differential term in the equations of motion must come from the laws of the elements.

EXAMPLE

Suppose it is desired to write the equations of motion of the dynamic network shown in Fig. 4-15. The capacitor is nonlinear and is characterized by a v-q curve which can be represented mathematically by $q = 2 - 3v^3 + 5v^5$. The inductor is also nonlinear and is characterized by an i-φ curve which can be rep-

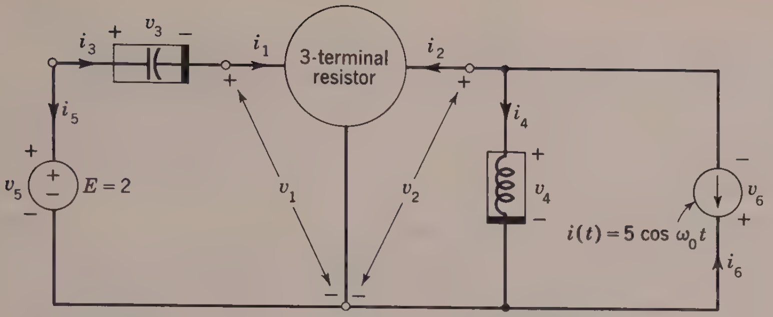


Fig. 4-15. A typical dynamic nonlinear network.

resented mathematically by $\varphi = 1 + 2i - 3i^2 + i^3$. The three-terminal resistor is characterized by two sets of curves which can be described mathematically by $i_1 = 1 + v_1 + 3v_1i_2^3 - 4i_2^5$ and $v_2 = 4 - i_2v_1 - 2i_2^2v_1^5 + v_1^3$. Element 5 is a 2-volt battery, and Element 6 is a current source with terminal current $i(t)$ equal to $5 \cos \omega_0 t$. The units of the voltages, currents, charges, and flux linkages are assumed to be in volts, amperes, coulombs, and webers, respectively.

For comparison purposes, the topology of this network is chosen to be identical with that of the resistive network shown in Fig. 4-9. Therefore, the equations from the laws of interconnection are given by Eqs. (4-18*a*) to (4-18*f*). To obtain the equations from the laws of elements of a dynamic network, we must realize that each capacitor introduces an additional variable, its charge q. Similarly, each inductor introduces an additional variable, its flux linkage φ. The laws of elements for each capacitor must consist of two independent equations, an equation representing the v-q curve and an equation relating the charge and current variables, namely, $i = dq/dt$. Similarly, the two equations associated with each inductor consist of an equation representing the i-φ curve and an equation relating the flux linkage and voltage variables, namely, $v = d\varphi/dt$.

Equations from the laws of elements

Three-terminal resistor:
$$\begin{cases} i_1 = 1 + v_1 + 3v_1i_2^3 - 4i_2^5 & \text{(4-44}a\text{)} \\ v_2 = 4 - i_2v_1 - 2i_2^2v_1^5 + v_1^3 & \text{(4-44}b\text{)} \end{cases}$$

Capacitor:
$$\begin{cases} q_3 = 2 - 3v_3^3 + 5v_3^5 & \text{(4-44}c\text{)} \\ i_3 = \dfrac{dq_3}{dt} = \dfrac{dq_3}{dv_3}\dfrac{dv_3}{dt} = (-9v_3^2 + 25v_3^4)\dfrac{dv_3}{dt} & \text{(4-44}d\text{)} \end{cases}$$

$$\text{Inductor:}\quad \begin{cases} \varphi_4 = 1 + 2i_4 - 3i_4^2 + i_4^3 & \text{(4-44}e\text{)} \\ v_4 = \dfrac{d\varphi_4}{dt} = \dfrac{d\varphi_4}{di_4}\dfrac{di_4}{dt} = (2 - 6i_4 + 3i_4^2)\dfrac{di_4}{dt} & \text{(4-44}f\text{)} \end{cases}$$

Element 5: $v_5 = 2$ (4-44*g*)

Element 6: $i_6 = -5 \cos \omega_0 t$ (4-44*h*)

These equations together with the six equations from the laws of interconnection (Eq. 4-18) constitute a system of 14 equations in the 14 variables; namely, $v_1, v_2, \ldots, v_6, i_1, i_2, \ldots, i_6, q_3$, and φ_4. Just as in the resistive case, we can immediately eliminate at least half of the variables by substituting the expressions in Eq. (4-44) in place of the corresponding variables in Eq. (4-18); thus

$$2 - v_1 - v_3 = 0 \tag{4-45a}$$

$$(4 - i_2v_1 - 2i_2^2v_1^5 + v_1^3) - (2 - 6i_4 + 3i_4^2)\frac{di_4}{dt} = 0 \tag{4-45b}$$

$$(2 - 6i_4 + 3i_4^2)\frac{di_4}{dt} + v_6 = 0 \tag{4-45c}$$

$$(-9v_3^2 + 25v_3^4)\frac{dv_3}{dt} + i_5 = 0 \tag{4-45d}$$

$$(1 + v_1 + 3v_1i_2^3 - 4i_2^5) - (-9v_3^2 + 25v_3^4)\frac{dv_3}{dt} = 0 \tag{4-45e}$$

$$i_2 + i_4 + 5 \cos \omega_0 t = 0 \tag{4-45f}$$

The above represents a system of six independent functional-differential equations; some are algebraic equations, and some are differential equations. Since no general mathematical methods exist for solving this class of equations, we must reduce them into a form which can be handled by existing techniques. The first step is to eliminate the functional equations in order to obtain a system of differential equations; that is, each equation must contain at least one derivative term. This can be done by solving for v_1 from Eq. (4-45*a*) and i_2 from Eq. (4-45*f*); thus

$$v_1 = 2 - v_3 \tag{4-46a}$$

$$i_2 = -5 \cos \omega_0 t - i_4 \tag{4-46b}$$

Substituting these in place of v_1 and i_2 in the remaining equations, we obtain

$$\begin{aligned}[4 - (-5 \cos \omega_0 t - i_4)(2 - v_3) - 2(-5 \cos \omega_0 t - i_4)^2(2 - v_3)^5 \\ + (2 - v_3)^3] - (2 - 6i_4 + 3i_4^2)\frac{di_4}{dt} = 0\end{aligned} \tag{4-47a}$$

$$(2 - 6i_4 + 3i_4^2)\frac{di_4}{dt} + v_6 = 0 \tag{4-47b}$$

$$(-9v_3^2 + 25v_3^4)\frac{dv_3}{dt} + i_5 = 0 \tag{4-47c}$$

$$[1 + (2 - v_3) + 3(2 - v_3)(-5\cos\omega_0 t - i_4)^3 - 4(-5\cos\omega_0 t - i_4)^5] - (-9v_3^2 + 25v_3^4)\frac{dv_3}{dt} = 0 \tag{4-47d}$$

The above represent four independent differential equations in the four unknowns v_3, v_6, i_4, and i_5. For this particular example, we observe that only the variables i_4 and v_3 appear in Eqs. (4-47*a*) and (*d*).[1] Therefore, these two equations can be solved independently of the remaining equations. Assuming, for the time being, that the solutions are given by $v_3 = f_3(t)$ and $i_4 = f_4(t)$, then the solutions for v_6 and i_5 can be found from Eqs. (4-47*b*) and (*c*) by direct substitution; thus

$$v_6(t) = -\{2 - 6f_4(t) + 3[f_4(t)]^2\}\frac{df_4(t)}{dt} \tag{4-48a}$$

$$i_5(t) = -\{-9[f_3(t)]^2 + 25[f_3(t)]^4\}\frac{df_3(t)}{dt} \tag{4-48b}$$

Therefore, the crucial step consists of solving for v_3 and i_4 from Eqs. (4-47*a*) and (*d*). To do this, let us rewrite these equations in the form

$$\frac{di_4}{dt} = \frac{4 - (-5\cos\omega_0 t - i_4)(2 - v_3) - 2(-5\cos\omega_0 t - i_4)^2(2 - v_3)^5 + (2 - v_3)^3}{2 - 6i_4 + 3i_4^2} \tag{4-49a}$$

$$\frac{dv_3}{dt} = \frac{1 + (2 - v_3) + 3(2 - v_3)(-5\cos\omega_0 t - i_4)^3 - 4(-5\cos\omega_0 t - i_4)^5}{-9v_3^2 + 25v_3^4} \tag{4-49b}$$

Upon examination of these equations we can make three important observations: (1) the term on the left-hand side of each equation is a first-order time derivative of a branch variable; (2) no time derivative terms appear on the right-hand side; (3) each dependent variable on the right-hand side appears also on the left-hand side of one of these equations. Any system of equations which satisfies these three properties is said to be in the *normal form,* and the independent variables in the equations are called *state variables.* The reason for reducing the equations of motion to the normal form lies in the fact that almost all techniques available for solving nonlinear differential equations have been enunciated in terms of the normal form. The techniques for solving this class of equations are discussed in Sec. 4-7. Accordingly, Eq. (4-49) will be solved in that section.

---

[1] In the more general case, the variables may appear in all equations.

4-6-1 SELECTION OF STATE VARIABLES OF NORMAL-FORM EQUATIONS

Since the most crucial step in the analysis of dynamic nonlinear networks is the reduction of the equations of motion into the normal form, let us study this topic in more detail.[1] By definition, the following system of first-order differential equations

$$\begin{aligned}\frac{dx_1}{dt} &= f_1(x_1,x_2,\ldots,x_n,t)\\ \frac{dx_2}{dt} &= f_2(x_1,x_2,\ldots,x_n,t)\\ &\cdots\cdots\cdots\cdots\cdots\cdots\\ \frac{dx_n}{dt} &= f_n(x_1,x_2,\ldots,x_n,t)\end{aligned} \tag{4-50}$$

is said to be in the normal form because (1) only first-order time derivatives of the state variables $x_1, x_2, \ldots, x_n$ appear on the left-hand side; (2) no time derivative terms appear on the right-hand side; and (3) the dependent variables coincide with the state variables that appear on the left-hand side. It is not always easy to reduce a system of functional-differential equations into the normal form. The basic problem lies in the choice of an appropriate set of state variables. To demonstrate the difficulty involved, let us examine Eq. (4-44*d*) more closely. Observe that to express i_3 in terms of v_3, it is necessary to calculate dq_3/dv_3. No problem was encountered earlier because the capacitor v-q curve is voltage-controlled, and, hence, the derivative can be easily obtained. However, suppose the given v-q curve happens to be charge-controlled, say

$$v_3 = q_3{}^3 - q_3 \tag{4-51}$$

In this case, it is necessary that we express q_3 in terms of v_3 before dq_3/dv_3 can be calculated. Unfortunately, this is not possible because q_3 is a multivalued function of v_3. This is equivalent to

[1] A general discussion of this topic is beyond the scope of this book. For a rigorous treatment, the reader is referred to the book by F. F. Kuo and W. G. Magnuson, Jr., "Computer Oriented Circuit Design," Prentice-Hall Inc., Englewood Cliffs, N.J., 1969.

saying that the inverse function of Eq. (4-51) does not exist. In this case, the normal-form equations cannot be obtained, if we insist on v_3 as the state variable. There is, of course, no reason why we should insist on choosing only voltages and currents as the state variables. Any other set of variables $x_1, x_2, \ldots, x_n$ is just as valid, provided (1) the resulting equations can be expressed in the normal form and (2) the voltage and current solutions can be calculated once $x_1(t), x_2(t), \ldots, x_n(t)$ are found.

EXAMPLE

Consider the same network shown in Fig. 4-15. This time, let the capacitor be characterized by a charge-controlled curve $v = 1 - 12q + q^3$. Let the inductor be characterized by a flux-controlled curve $i = \varphi - \varphi^3$. All other elements remain unchanged. The reader should sketch these two curves and verify that their inverse functions do not exist. Therefore, we must rewrite Eq. (4-44) into the following form:

Equations from the laws of elements

Three-terminal resistor:
$$\begin{cases} i_1 = 1 + v_1 + 3v_1 i_2^3 - 4i_2^5 & \text{(4-52}a\text{)} \\ v_2 = 4 - i_2 v_1 - 2i_2^2 v_1^5 + v_1^3 & \text{(4-52}b\text{)} \end{cases}$$

Capacitor:
$$\begin{cases} v_3 = 1 - 12q_3 + q_3^3 & \text{(4-52}c\text{)} \\ i_3 = \dfrac{dq_3}{dt} & \text{(4-52}d\text{)} \end{cases}$$

Inductor:
$$\begin{cases} i_4 = \varphi_4 - \varphi_4^3 & \text{(4-52}e\text{)} \\ v_4 = \dfrac{d\varphi_4}{dt} & \text{(4-52}f\text{)} \end{cases}$$

Element 5: $v_5 = 2$ (4-52*g*)

Element 6: $i_6 = -5 \cos \omega_0 t$ (4-52*h*)

Substituting these expressions for the corresponding variables in the equations from the laws of interconnection [Eq. (4-18)], we obtain

$$2 - v_1 - (1 - 12q_3 + q_3^3) = 0 \tag{4-53a}$$

$$(4 - i_2 v_1 - 2i_2^2 v_1^5 + v_1^3) - \frac{d\varphi_4}{dt} = 0 \tag{4-53b}$$

$$\frac{d\varphi_4}{dt} + v_6 = 0 \tag{4-53c}$$

$$\frac{dq_3}{dt} + i_5 = 0 \tag{4-53d}$$

$$(1 + v_1 + 3v_1 i_2^3 - 4i_2^5) - \frac{dq_3}{dt} = 0 \tag{4-53e}$$

$$i_2 + (\varphi_4 - \varphi_4^3) + 5 \cos \omega_0 t = 0 \tag{4-53f}$$

As usual, the first step is to reduce these functional-differential equations into a system of differential equations. This can be done by solving for v_1 in terms of q_3 from Eq. (4-53*a*) and i_2 in terms of φ_4 from Eq. (4-53*f*); thus

$$v_1 = 1 + 12q_3 - q_3^3 \tag{4-54a}$$

$$i_2 = -5 \cos \omega_0 t - \varphi_4 + \varphi_4^3 \tag{4-54b}$$

Substituting these equations in place of v_1 and i_2 in the remaining equations, we obtain

$$\begin{aligned}[4 &- (-5 \cos \omega_0 t - \varphi_4 + \varphi_4^3)(1 + 12q_3 - q_3^3) \\ &\quad - 2(-5 \cos \omega_0 t - \varphi_4 + \varphi_4^3)^2(1 + 12q_3 - q_3^3)^5 \\ &\qquad + (1 + 12q_3 - q_3^3)^3] - \frac{d\varphi_4}{dt} = 0\end{aligned} \tag{4-55a}$$

$$\frac{d\varphi_4}{dt} + v_6 = 0 \tag{4-55b}$$

$$\frac{dq_3}{dt} + i_5 = 0 \tag{4-55c}$$

$$\begin{aligned}[1 &+ (1 + 12q_3 - q_3^3) \\ &\quad + 3(1 + 12q_3 - q_3^3)(-5 \cos \omega_0 t - \varphi_4 + \varphi_4^3)^3 \\ &\qquad - 4(-5 \cos \omega_0 t - \varphi_4 + \varphi_4^3)^5] - \frac{dq_3}{dt} = 0\end{aligned} \tag{4-55d}$$

Again, we observe that v_6 and i_5 can be calculated from Eqs. (4-55*b*) and (*c*), once φ_4 and q_3 are found. Our task, therefore, is to recast Eqs. (4-55*a*) and (*d*) into the normal form

$$\begin{aligned}\frac{d\varphi_4}{dt} &= 4 - (-5 \cos \omega_0 t - \varphi_4 + \varphi_4^3)(1 + 12q_3 - q_3^3) \\ &\quad - 2(-5 \cos \omega_0 t - \varphi_4 + \varphi_4^3)^2(1 + 12q_3 - q_3^3)^5 \\ &\qquad + (1 + 12q_3 - q_3^3)^3\end{aligned} \tag{4-56a}$$

$$\begin{aligned}\frac{dq_3}{dt} &= 2 + 12q_3 - q_3^3 + 3(1 + 12q_3 - q_3^3)(-5 \cos \omega_0 t - \varphi_4 + \varphi_4^3)^3 \\ &\qquad - 4(-5 \cos \omega_0 t - \varphi_4 + \varphi_4^3)^5\end{aligned} \tag{4-56b}$$

These equations can then be solved by the techniques to be described in Sec. 4-7.

4-6-2 AUTONOMOUS AND NONAUTONOMOUS NETWORKS

The independent time variable t normally appears in the right-hand side of the normal-form equations whenever the network contains one or more ac (time-varying) sources. For example, the variable t appears in both Eqs. (4-49) and (4-56) in the form of $5 \cos \omega_0 t$, the terminal current of the alternating current source. This type of normal-form equation is said to be *nonautonomous,* and its associated network is said to be a *nonautonomous network.*

If a dynamic network does not contain any ac (time-varying) source, then the normal-form equations will not contain the variable t on the right-hand side; that is,

$$\begin{aligned}\frac{dx_1}{dt} &= f_1(x_1,x_2, \ldots ,x_n)\\ \frac{dx_2}{dt} &= f_2(x_1,x_2, \ldots ,x_n)\\ &\cdots\cdots\cdots\cdots\cdots\cdots\\ \frac{dx_n}{dt} &= f_n(x_1,x_2, \ldots ,x_n)\end{aligned} \tag{4-57}$$

This type of equation is said to be an *autonomous* system of equations, and its associated network is said to be an *autonomous network.* The reason for distinguishing between these two types of dynamic networks is that their methods of solution are generally quite different.

Exercise 1: Identify which of the following equations are in the normal form, and classify each according to whether they are autonomous or nonautonomous.

(*a*) $\frac{dv_1}{dt} = v_1v_3 + 2 \sin t \qquad \frac{dv_3}{dt} = -v_1^3 + v_3^2$

(*b*) $\frac{di_1}{dt} = 2i_1 + v_1^2 + v_2 \qquad \frac{dv_2}{dt} = i_1v_2 + 10v_2$

(*c*) $q_1\frac{dq_1}{dt} = 2q_1 + q_2^3 \qquad \frac{dq_2}{dt} = q_1q_2 + q_1^2$

(*d*) $\frac{di_1}{dt} = q_2^3 \qquad \frac{dq_2}{dt} = v_3 \qquad \frac{dv_3}{dt} = i_1^2$

Exercise 2: Show that any second-order differential equation of the form $d^2y/dt^2 = f(y,dy/dt)$ can always be transformed into the normal form by defining $x_1 = y$ and $x_2 = dy/dt$.

Exercise 3: Find the normal-form equations of the network shown in Fig. 4-15 but with the capacitor and the inductor characterized, respectively, by $q = v - v^3$ and $i = \varphi^3 - \varphi^5$.

Exercise 4: Show that by defining a set of n state variables $x_1 = y$, $x_2 = dy/dt$, $\ldots$, $x_n = d^{n-1}y/dt^{n-1}$, any nth-order differential equation of the form $d^n y/dt^n = f(y, dy/dt, d^2y/dt^2, \ldots, d^{n-1}y/dt^{n-1})$ can be transformed to the normal form.

4-7 PRACTICAL METHODS FOR SOLVING NONLINEAR DIFFERENTIAL EQUATIONS

The analysis of dynamic nonlinear networks generally entails the solution of a system of nonlinear differential equations. This is usually a formidable task unless the equations are cast in the normal form. Accordingly, our discussion will be restricted to this class of differential equations.

Four general types of methods are available for solving nonlinear differential equations in the normal form, namely, *numerical, graphical, piecewise-linear,* and *analog simulation methods.* We have already seen an example of an analog simulation method in Chap. 3. As the name implies, this method involves the simulation of the equations by means of an analog computer. It is not a very general method for nonlinear systems, but it is useful for solving many practical problems. Since the analog method is normally taught as a separate course, we shall not discuss it any further. The piecewise-linear and the graphical methods are particularly useful for solving systems consisting of either one or two equations. The advantage of these methods is that they do not require much mathematical background. Accordingly, Chaps. 14 to 17 are devoted to piecewise-linear methods, and Chaps. 18 and 19 are devoted to graphical methods. The numerical methods are ideally suited for digital computers and are widely used in practice. Most of these methods require some background in differential equations and are therefore beyond the scope of this book. However, there is one numerical method which has a very simple geometrical interpretation, namely, the Euler algorithm.

4-7-1 EULER ALGORITHM FOR ONE DIFFERENTIAL EQUATION

Consider the simplest case of a nonlinear differential equation

$$\frac{dx}{dt} = f(x,t) \tag{4-58}$$

A time function $x = X(t)$ is said to be a solution of Eq. (4-58) if

$$\frac{dX(t)}{dt} \equiv f(X(t),t)$$

For example, if $dx/dt = x$, then $x = ke^t$ is a solution of this differential equation for any value of k. In order to specify a numerical value for the arbitrary constant k, we must specify the *initial condition* $x = x_0$ at $t = t_0$ that must be satisfied by the solution. For example, if $x_0 = 10$ at $t_0 = 0$, then $10 = ke^0 = k$, and the solution is therefore given by $x = 10e^t$. The above example is a rather simple one. In most nonlinear problems it is impossible to find an exact analytical solution; then we must resort to approximate solutions.

The simplest approximate method for finding the solution of Eq. (4-58) with an initial condition $x(t_0) = x_0$ is obtained by writing Eq. (4-58) in the limiting form

$$\lim_{\Delta t \to 0} \frac{\Delta x}{\Delta t} = f(x,t) \tag{4-59}$$

If we restrict Δt to be a relatively small number but not zero, then Eq. (4-59) can be written approximately as

$$\Delta x \approx hf(x,t) \tag{4-60}$$

where $h \equiv \Delta t$ is called the *step size*. Hence, if $x(t) = x_0$ at $t = t_0$, then at $t = t_0 + h$,

$$\Delta x \approx hf(x_0,t_0)$$

The approximate value of $x(t)$ at $t_1 \equiv t_0 + h$ is, therefore, given by

$$x(t_1) \approx x_0 + hf(x_0,t_0) \equiv x_1$$

We can obviously repeat this procedure to obtain

$$x(t_2) \approx x_1 + hf(x_1,t_1) \equiv x_2$$
$$x(t_3) \approx x_2 + hf(x_2,t_2) \equiv x_3$$
$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$
$$x(t_n) \approx x_{n-1} + hf(x_{n-1},t_{n-1}) \equiv x_n$$

where $t_2 = t_1 + h$, $t_3 = t_2 + h, \ldots, t_n = t_{n-1} + h$. If the solution is desired over some time interval $t_0 \leq t \leq T$, then, corresponding to a chosen step size h, a total of n iterations is necessary, where n is chosen such that $t_n \leq T$ but $t_n + h > T$. Hence, corresponding

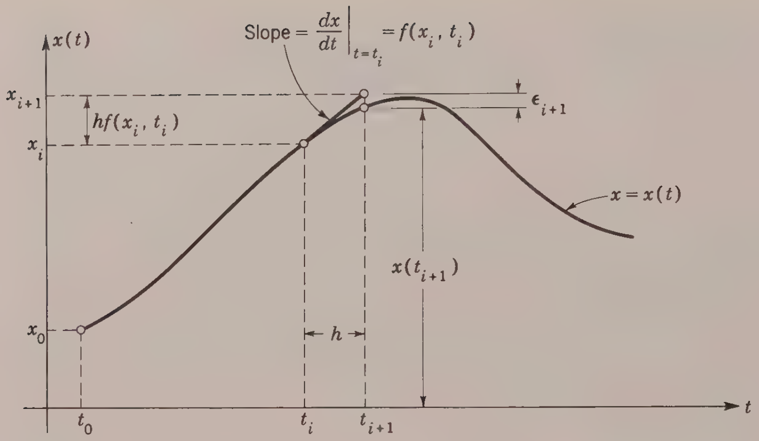


Fig. 4-16. The geometrical interpretation of the Euler algorithm.

to the initial condition $x(t_0) = x_0$, the above algorithm can be summarized as follows:

$$\begin{aligned} x_{i+1} &= x_i + hf(x_i,t_i) \\ t_{i+1} &= t_i + h \qquad\qquad (4\text{-}61) \\ i &= 0, 1, 2, \ldots, n-1 \end{aligned}$$

This procedure is generally known as the *Euler algorithm.* A simple geometrical interpretation of this algorithm is shown in Fig. 4-16. Observe that the distance $x_{i+1} - x_i$ is equal to the product of the step size h and the slope $dx/dt = f(x,t)$ at $t = t_i$. The error of approximation

$$\epsilon_{i+1} \equiv x_{i+1} - x(t_{i+1}) \qquad (4\text{-}62)$$

is called the *truncation error.* Clearly, if the solution $x(t)$ is a relatively smooth curve, and if the step size h is relatively small, then the approximation will be quite reasonable. Of course, a small step size would require a large number of iterations n. In practice, the iteration is invariably done by a digital computer, as are most other numerical techniques.

As a simple example, consider the differential equation

$$\frac{dx}{dt} = 1 + x^2 \qquad (4\text{-}63)$$

and the initial condition

$$x(0) = 0 \qquad (4\text{-}64)$$

Assuming $h = 0.1$ and $n = 10$, the Euler algorithm is given by

$$x_{i+1} = x_i + 0.1(1 + x_i^2)$$
$$t_{i+1} = t_i + 0.1 \tag{4-65}$$
$$i = 0, 1, 2, \ldots, 9$$

The result of this iteration is tabulated on the first two columns of Table 4-3. The above example happens to have an exact analytical solution

$$x(t) = \tan t \tag{4-66}$$

For comparison purposes, Eq. (4-66) is calculated at the same time increments, and the result is tabulated in the third column. The truncation error ϵ_i corresponding to each time t_i is shown on the last column.

4-7-2 EULER ALGORITHM FOR TWO OR MORE DIFFERENTIAL EQUATIONS

The same procedure can be extended to solve a system of n differential equations in the normal form. Since the algorithm is similar in all cases, we shall consider only a system of two differential equations

$$\frac{dx}{dt} = f_1(x,y,t)$$
$$\frac{dy}{dt} = f_2(x,y,t) \tag{4-67}$$

TABLE 4-3 *A comparison between the solution obtained numerically by the Euler algorithm and the exact solution $x = \tan t$.*

TIME T	SOLUTION BY EULER ALGORITHM	EXACT SOLUTION X=TAN T	TRUNCATION ERROR
0.00	0.0000	0.0000	0.0000
0.10	0.1000	0.1003	-0.0003
0.20	0.2010	0.2027	-0.0017
0.30	0.3050	0.3093	-0.0043
0.40	0.4143	0.4228	-0.0084
0.50	0.5315	0.5463	-0.0148
0.60	0.6598	0.6841	-0.0244
0.70	0.8033	0.8423	-0.0390
0.80	0.9678	1.0296	-0.0618
0.90	1.1615	1.2602	-0.0987
1.00	1.3964	1.5574	-0.1610

TABLE 4-4 *Computer-printed output of the solutions of Eq. (4-49) by the Euler algorithm.*

T	X	Y	T	X	Y	T	X	Y
0.00	-2.0000	1.5000	0.00	-3.0000	2.5000	0.00	-4.5000	3.5000
0.10	-1.9805	2.3776	0.10	-2.9934	2.5153	0.10	-4.4959	3.5000
0.20	-1.9690	2.5105	0.20	-2.9867	2.5295	0.20	-4.4921	3.5000
0.30	-1.9574	2.6071	0.30	-2.9799	2.5414	0.30	-4.4893	3.5000
0.40	-1.9440	2.6760	0.40	-2.9731	2.5500	0.40	-4.4877	3.5000
0.50	-1.9284	2.7231	0.50	-2.9662	2.5554	0.50	-4.4869	3.5000
0.60	-1.9113	2.7531	0.60	-2.9592	2.5582	0.60	-4.4859	3.5000
0.70	-1.8938	2.7705	0.70	-2.9520	2.5594	0.70	-4.4824	3.4999
0.80	-1.8768	2.7793	0.80	-2.9445	2.5597	0.80	-4.4732	3.4999
0.90	-1.8611	2.7830	0.90	-2.9367	2.5598	0.90	-4.4541	3.4996
1.00	-1.8464	2.7842	1.00	-2.9285	2.5599	1.00	-4.4200	3.4988
1.10	-1.8323	2.7845	1.10	-2.9198	2.5599	1.10	-4.3650	3.4966
1.20	-1.8181	2.7845	1.20	-2.9104	2.5599	1.20	-4.2832	3.4909
1.30	-1.8026	2.7845	1.30	-2.9002	2.5590	1.30	-4.1691	3.4788
1.40	-1.7849	2.7845	1.40	-2.8891	2.5548	1.40	-4.0196	3.4555
1.50	-1.7634	2.7842	1.50	-2.8769	2.5394	1.50	-3.8363	3.4144
1.60	-1.7369	2.7822	1.60	-2.8638	2.4951	1.60	-3.6289	3.3467
1.70	-1.7040	2.7745	1.70	-2.8503	2.3825	1.70	-3.4189	3.2385
1.80	-1.6638	2.7521	1.80	-2.8377	2.0998	1.80	-3.2394	3.0623
1.90	-1.6172	2.6982	1.90	-2.8275	1.2080	1.90	-3.1275	2.7463
2.00	-1.5692	2.5803	2.00	-2.8539	-15.5571	2.00	-3.0925	2.0190

with the initial conditions $x(t_0) = x_0$ and $y(t_0) = y_0$. The Euler algorithm in this case is given by

$$
\begin{aligned}
x_{i+1} &= x_i + hf_1(x_i,y_i,t_i)\\
y_{i+1} &= y_i + hf_2(x_i,y_i,t_i)\\
t_i &= t_i + h\\
i &= 0, 1, 2, \ldots, n-1
\end{aligned}
\tag{4-68}
$$

Let us now apply this algorithm to find the approximate solutions $i_4(t)$ and $v_3(t)$ of Eq. (4-49). Suppose the solution is desired over the time interval $0 \leq t \leq 2.0$. If a step size $h = 0.1$ is chosen, then the value of n is equal to 20. The algorithm for this case is therefore given by[1]

$$
\begin{aligned}
&x_{i+1}\\
&=x_i+0.1\left[\frac{4-(-5\cos t_i-x_i)(2-y_i)-2(-5\cos t_i-x_i)^2(2-y_i)^5+(2-y_i)^3}{(2-6x_i+3x_i^2)}\right]\\
&y_{i+1}=y_i+0.1\left[\frac{1+(2-y_i)+3(2-y_i)(-5\cos t_i-x_i)^3-4(-5\cos t_i-x_i)^5}{(-9y_i^2+25y_i^4)}\right]\\
&t_{i+1}=t_i+0.1\\
&i=0, 1, 2, \ldots, 19
\end{aligned}
\tag{4-69}
$$

The results of carrying out the above iterations for three different sets of initial conditions are given in Table 4-4. The reader should recognize that this formidable computational task can be easily carried out with the help of a digital computer.

---

[1] For simplicity, we have assumed $\omega_0 = 1$ in Eq. (4-49).

Exercise 1: Use the Euler algorithm to calculate the approximate solution of $dx/dt = x$ over the time interval $0 < t < 1$. Assume an initial condition $x(0) = 10$ and a step size $h = 0.2$. Compare the solution with the exact answer $x(t) = 10e^t$.

Exercise 2: Specify the Euler algorithm for a system of three differential equations in the normal form.

4-8 PRINCIPLES OF DUALITY

There are many physical phenomena or systems in nature which have occurred in *dual* forms. Generally speaking, we say two systems or phenomena are *duals* of each other if we can exhibit some kind of one-to-one correspondence between various quantities or attributes of the two systems. For example, in mathematics, two equations which differ only in symbols but are otherwise identical in form are said to be dual equations. In physics, for each translational system or problem there exists a corresponding rotational system or problem, and they are usually referred to as dual systems or problems. The recognition of dual quantities, attributes, phenomena, properties, or concepts often leads to the discovery and invention of new ideas. In electrical engineering, the application of the principle of duality has often led not only to a simplification of solutions but also to the discovery and invention of new useful networks.

Before we render the concepts of duality more precise, it is instructive to consider first the two nonlinear networks shown in Fig. 4-17*a* and *b*. The equations of motion of these two networks are readily obtained and are tabulated in Table 4-5. A careful comparison of the expressions in the two columns of this table re-

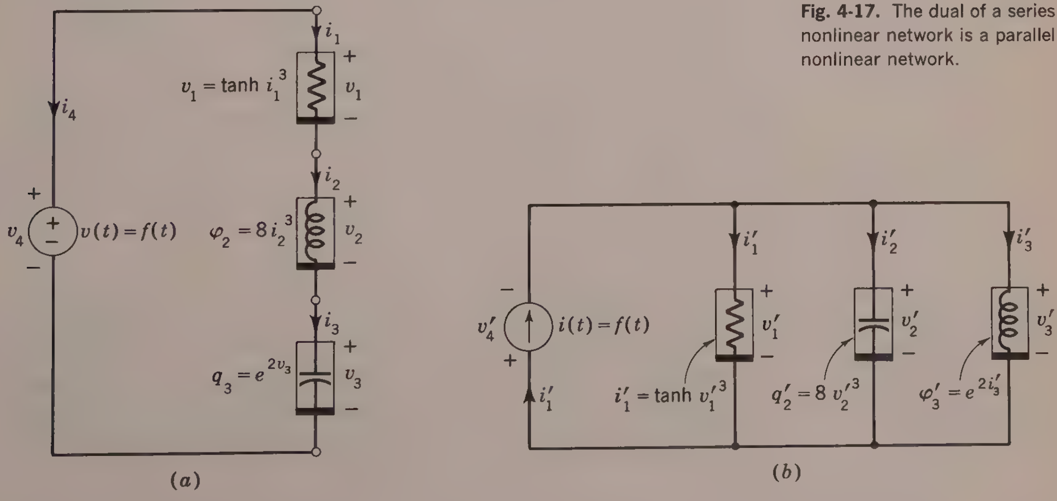


Fig. 4-17. The dual of a series nonlinear network is a parallel nonlinear network.

TABLE 4-5 *Equations of motion for the networks in Fig. 4-17a and b.*

Network of Fig. 4-17*a*	Network of Fig. 4.17*b*
Laws of Elements	Laws of Elements
$v_1 = \tanh i_1{}^3$	$i'_1 = \tanh v'_1{}^3$
$v_2 = \dfrac{d\varphi_2}{dt} = \dfrac{d\varphi_2}{di_2}\dfrac{di_2}{dt} = 24 i_2{}^2 \dfrac{di_2}{dt}$	$i'_2 = \dfrac{dq'_2}{dt} = \dfrac{dq'_2}{dv'_2}\dfrac{dv'_2}{dt} = 24 v'_2{}^2 \dfrac{dv'_2}{dt}$
$i_3 = \dfrac{dq_3}{dt} = \dfrac{dq_3}{dv_3}\dfrac{dv_3}{dt} = 2e^{2v_3}\dfrac{dv_3}{dt}$	$v'_3 = \dfrac{d\varphi'_3}{dt} = \dfrac{d\varphi'_3}{di'_3}\dfrac{di'_3}{dt} = 2e^{2i'_3}\dfrac{di'_3}{dt}$
$v_4 = f(t)$	$i'_4 = f(t)$
Laws of Interconnection	Laws of Interconnection
KVL: $v_1 + v_2 + v_3 - v_4 = 0$	KCL: $i'_1 + i'_2 + i'_3 - i'_4 = 0$
KCL: $i_1 + i_4 = 0$	KVL: $v'_1 + v'_4 = 0$
$i_1 - i_2 = 0$	$v'_1 - v'_2 = 0$
$i_2 - i_3 = 0$	$v'_2 - v'_3 = 0$

veals a one-to-one correspondence between the equations. As a matter of fact, except for the symbols, the equations in the two columns are identical in form. Observe that, had we replaced v_j by i'_j, i_j by v'_j, φ_j by q'_j, and q_j by φ'_j for the variables in the left column, the result would be identical with that in the right column, and therefore the two networks are said to be dual networks. One of the significant facts about dual networks is that once we know the solution of one network, the solution of the dual network can be obtained immediately by simply interchanging the symbols. This means that as soon as we know the behavior and properties of one network, we immediately know the behavior and properties of the dual network. Hence a lot of redundancy is avoided if we can recognize dual networks. In this book, there will be many occasions when we shall take advantage of the principle of duality. In view of its importance, we shall now precisely define the concept of duality.

DEFINITION OF DUAL NETWORKS

Let N and N' be a pair of networks each containing b two-terminal network elements which are not controlled sources.[1] Then N and N' are dual networks if the elements in N and N' can be labeled, respectively, as $b_1, b_2, \ldots, b_b$ and $b'_1, b'_2, \ldots, b'_b$ such that the equations of motion of the two networks are identical in form.

---

[1] It is possible to generalize the definition of dual networks to include controlled sources. However, the procedure for constructing such networks is more complicated.

That is, if we replace v_j by i_j', i_j by v_j', φ_j by q_j', and q_j by φ_j' in the equations of motion for network N, we obtain the equations of motion for network N', and vice versa.

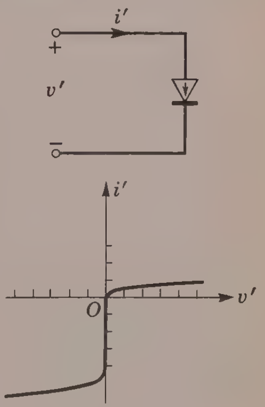
In view of the above definition, it is clear that the two networks N and N' are dual networks if the solution v_j, i_j, φ_j, or q_j for element b_j in network N is equal to the solution i_j', v_j', q_j', or φ_j' in element b_j' of N', and vice versa. Our next task will be to investigate what classes of networks would satisfy the above definition. This question is clearly equivalent to the following problem: Given a network N containing two-terminal elements, does there exist another network N' which is the dual of N, and if so, how can we find N'? To answer this question, it is necessary to uncover the duality relationships that must be satisfied by the laws of elements and the laws of interconnection of the two dual networks.

4-8-1 DUALITY RELATIONSHIPS FROM THE LAWS OF ELEMENTS

The definition for dual networks tells us that the equation of motion for element b_j is the dual of the equation of motion for element b_j' if the following one-to-one correspondence exists between the elements in N and N'.

Resistor If element b_j is a two-terminal resistor in N characterized by a curve Γ in the v-i plane, then the corresponding dual element b_j' in N' must be also a two-terminal resistor characterized by the same curve Γ in the i'-v' plane, i.e., with the i axis replaced by the v' axis and the v axis replaced by the i' axis. For example, if element b_j of N is a resistor characterized by $i_j = v_j^3 - 3v_j$, then the dual resistor in N' is a resistor characterized by $v_j' = i_j'^3 - 3i_j'$. In particular, if element b_j in N is a voltage source with $v_s(t) = K \sin t$, then the dual element b_j' in N' must be a current source with $i_s'(t) = K \sin t$. Observe that the dual of a given resistor is a new resistor, which may need a new name and a new symbol. For example, the dual of a zener diode with a v-i curve as defined in Table 1-1 is a new resistor whose new symbol and v'-i' curve are shown in Fig. 4-18. However, there are some two-terminal elements which have the interesting property that the dual of the element is the same element with its two terminals interchanged. For such elements, a new symbol is not needed for we only have to interchange the terminals of the given element to obtain its dual. The simplest example of this type of element is the

Fig. 4-18. The symbol and the v-i curve of the dual of a zener diode.



ideal diode. Clearly, the dual of an ideal diode is obtained simply by reversing the two terminals of the diode, as shown in Fig. 4-19. Observe that in contrast with the new symbol introduced in Fig. 4-18, no new symbol is needed to specify unambiguously the dual of an ideal diode.

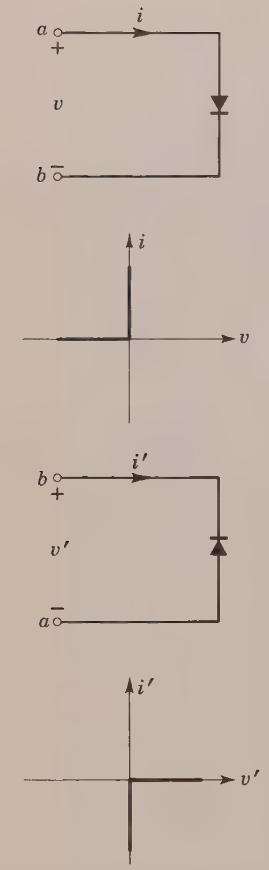


Fig. 4-19. The dual of an ideal diode is also an ideal diode with its terminals interchanged.

Inductor If element b_j in N is a two-terminal inductor characterized by a curve Γ in the i-φ plane, then the corresponding dual element b_j' in N' must be a capacitor characterized by the same curve Γ in the v'-q' plane; that is, simply change the variables φ and i to q' and v' while retaining the curve Γ. For example, the dual of an inductor characterized by $\varphi = \log i$ is a capacitor characterized by $q' = \log v'$.

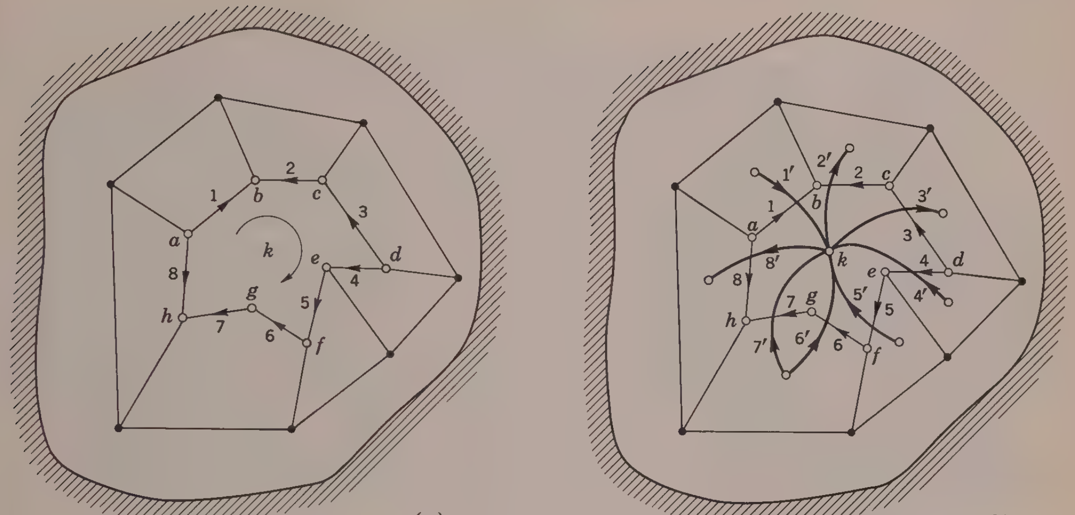
Capacitor If element b_j in N is a two-terminal capacitor characterized by a curve Γ in the v-q plane, then the corresponding dual element b_j' in N' must be an inductor characterized by the same curve Γ in the i'-φ' plane; that is, simply change the variables q and v to φ' and i' while retaining the curve Γ. For example, the dual of a capacitor characterized by $q = \tanh v$ is an inductor characterized by $\varphi' = \tanh i'$.

4-8-2 DUALITY RELATIONSHIPS FROM THE LAWS OF INTERCONNECTION

Since the equations of motion from the laws of interconnection are independent of the network elements and can be written directly from the network graph, it is clear that the duality relationships for these equations can be found directly from the network graphs of the dual networks. Now, from the definition of dual networks, each branch voltage v_j in N becomes a branch current i_j' in N', and each branch current i_j in N becomes a branch voltage v_j' in N'. It is clear that the equations of motion due to KVL in N must correspond to the equations of motion due to KCL in N', and vice versa. This means that if the network N has n nodes and b branches, resulting in $(n - 1)$ independent KCL equations and $b - (n - 1)$ independent KVL equations,[1] then the dual network N' must contain $(n - 1)$ *fundamental loops,* or meshes, and $b - (n - 1) + 1$ nodes in order to have correspondingly $(n - 1)$ independent KVL equations and $b - (n - 1)$ KCL equations. For example, the network N shown in Fig. 4-17*a* has four nodes and four branches, and the dual network shown in Fig. 4-17*b* has, correspondingly, three meshes, as expected. Given a network graph N with n nodes and

---

[1] Review the independent KVL and KCL equations criteria in Sec. 4-3.



(a)

(b)

Fig. 4-20. A typical mesh k of a planar network and its dual branches.

b branches, we shall now attempt to find a procedure (if it exists) to construct a dual network graph N' which we know must consist of $(n-1)$ fundamental loops, or meshes, and b branches.

If the network is planar, we know from the KVL mesh equation criteria (Sec. 4-3) that the number of independent KVL equations is equal to the number of meshes, namely, $b-(n-1)$. Hence, the dual network N' must have $b-(n-1)$ independent KCL equations. This means that N' must contain $b-(n-1)+1$ nodes. Let us consider the typical planar graph shown in Fig. 4-20*a*. In particular, consider a typical mesh such as mesh k formed by the branches 1, 2, 3, 4, 5, 6, 7, and 8. The KVL equation around mesh k is given by[1]

$$v_1 - v_2 - v_3 + v_4 + v_5 + v_6 + v_7 - v_8 = 0 \tag{4-70}$$

Now if N' is the dual of N, then the graph of N' must necessarily contain a corresponding node k whose KCL equation is the dual of Eq. (4-70), namely,

$$i_1' - i_2' - i_3' + i_4' + i_5' + i_6' + i_7' - i_8' = 0 \tag{4-71}$$

We can accomplish the above task if we place node k inside mesh k and, for each branch b_j around mesh k, draw a corresponding branch b_j' from node k to a node placed inside the mesh having branch b_j in common with mesh k, as shown in Fig. 4-20*b*. In view

---

[1] Recall that when we draw the network graph, we automatically assume that the positive polarity of each two-terminal element is located at the tail end of the current reference direction. This assumption must always be kept in mind. In particular, before we find the dual of a two-terminal element, we must first see to it that the characteristic curve of the element is specified with a set of references consistent with the above assumption.

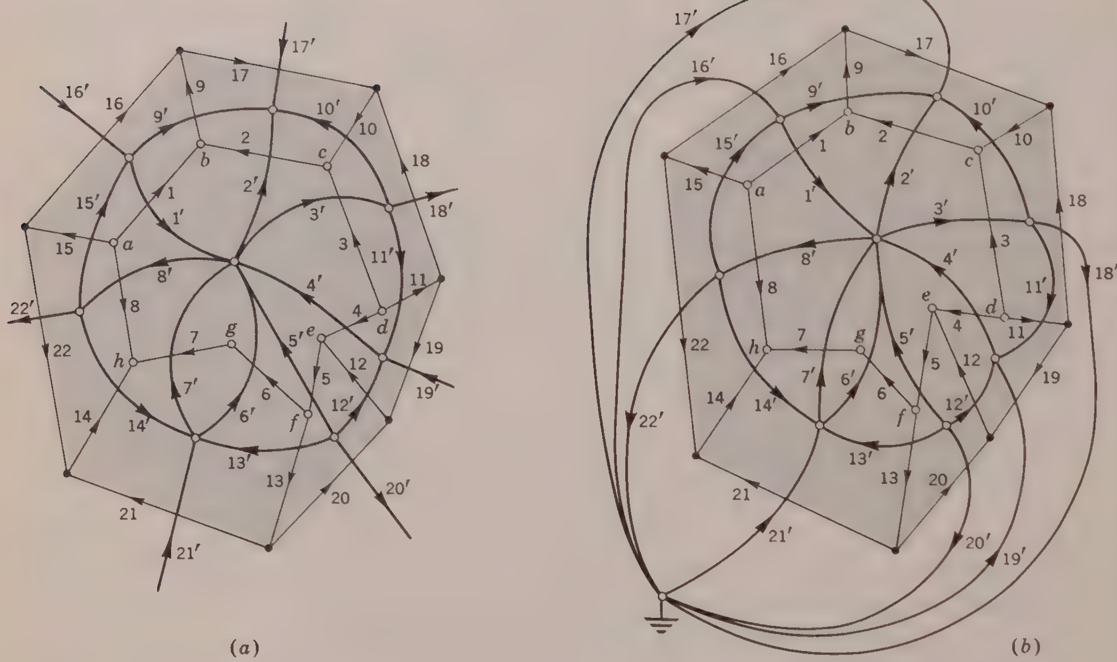


Fig. 4-21. The procedure for constructing a dual graph is as follows: (*a*) put a node inside each mesh and a node outside the network; (*b*) corresponding to each branch b_j common to two meshes, connect a branch b_j' between the two nodes in these meshes such that it cuts the branch b_j; and (*c*) assign a direction for b_j' in accordance with the reference connection.

of the above construction procedure, each branch b_j' in the dual graph must necessarily intersect branch b_j of the given graph. Finally, let us assign reference directions to each branch b_j' in the dual graph according to the following convention.

REFERENCE CONVENTION FOR DUAL GRAPH

Branch b_j' is assigned a direction toward node k if the corresponding branch b_j of the given graph is in a clockwise direction with respect to node k. Otherwise, branch b_j' will be assigned a direction away from node k.

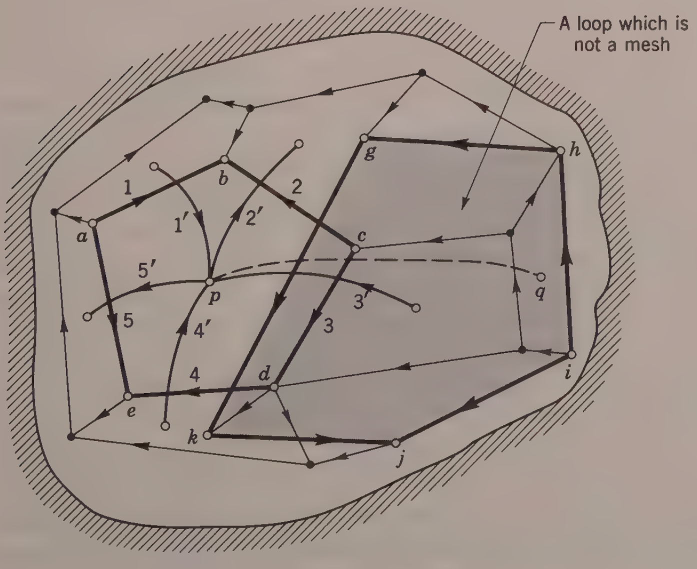
With the branches in Fig. 4-20*b* directed consistently according to the above convention, it is clear that we indeed obtain the desired Eq. (4-71).

If we repeat the above procedure for each of the $b - (n - 1)$ meshes, we would eventually end up with a network graph such as the examples shown in Fig. 4-21*a*, with branches cutting the outer boundary of the given graph N, and so far left unconnected. To complete the dual graph, we observe that KVL must be satisfied around the outer boundary of the given graph N, and similarly, in the dual graph, the branches cutting the outer boundary in Fig. 4-21*a* must satisfy KCL. This condition can be satisfied by placing one node outside the dual graph to terminate all branches of the dual graph which cut the outer boundary of N at the external node. The resulting network graph shown in Fig. 4-21*b* is then the dual of the given network N. This procedure, together with the principle for drawing dual elements, would enable us to find the dual of any planar network.

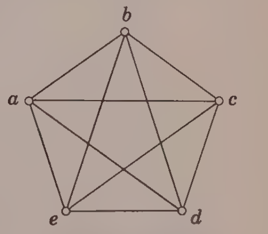
What happens if network N is nonplanar? Since the concept of a mesh is undefined for nonplanar networks, the dual quantities for this class of networks are loops and nodes. Moreover, since the network is nonplanar, there must be at least one closed loop that is not a mesh (remember that a mesh encloses no branches, a closed loop does), such as the typical case shown in Fig. 4-22*a*.

Consider first writing KVL around loop *abcde* consisting of branches 1, 2, 3, 4, and 5. Using the procedure for a planar network, we place a node p inside loop *abcde* with five branches as shown. This would give us the desired KCL equation at node p. However, suppose we repeat the procedure for loop *ghijk* and place a node q inside it. The procedure for drawing a planar network would now require a branch connecting nodes p and q,

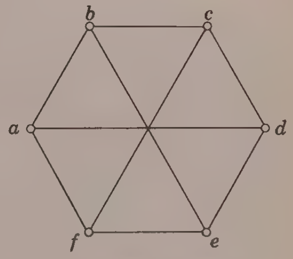
Fig. 4-22. The network shown in (*a*) demonstrates that a nonplanar network cannot possess a dual. A network graph is nonplanar if, and only if, it contains either one of the two Kuratowski subgraphs shown in (*b*) and (*c*).



(*a*)



(*b*) Kuratowski subgraph 1



(*c*) Kuratowski subgraph 2

thus adding one more branch to node p. But this would violate the desired KCL equation at node p. Hence the procedure would not work for a nonplanar network. Since the above argument is valid in general, it is impossible to find a network graph which is the dual of a nonplanar network. This result is important enough to be stated as a theorem.

DUAL NETWORK EXISTENCE THEOREM

A network N has a dual if, and only if, N is a planar network.

It is clear from this theorem that before we attempt to find the dual of a given network, we must determine whether it is planar. For simple networks, this can be readily determined by inspection. For more complicated networks, however, it is sometimes not easy to distinguish a planar from a nonplanar network. Kuratowski's theorem applies to this situation, but because it is extremely difficult to prove, it is only stated here.

KURATOWSKI'S THEOREM

The necessary and sufficient condition that a network N be planar is that the network graph contain neither the subgraph (i.e., part of the graph obtained by removing some branches and nodes) shown in Fig. 4-22*a* nor the subgraph shown in Fig. 4-22*b*.[1]

As a consequence of Kuratowski's theorem, it is clear that any network with less than five nodes is necessarily planar. Hence, without having to redraw the network in Fig. 4-4*a*, we can conclude that it is planar because it contains only four nodes. When a network contains more than five nodes, we can apply Kuratowski's theorem by searching for the existence of a set of five nodes with each node having a branch connected to each of the other nodes, as in Fig. 4-22*b*. Or we can look for the existence of a set of six nodes with branches connected in the form of Fig. 4-22*c*. For example, the network shown earlier in Fig. 4-5 is nonplanar because it contains a Kuratowski subgraph of the form shown in Fig. 4-22*b*.

It is easier to prove that a network is nonplanar because we need only exhibit one of the two Kuratowski subgraphs. To prove that a network is planar, we must examine all possible groups of branches and nodes to ascertain that the network does not contain either of the two Kuratowski subgraphs.

---

[1] This theorem implicitly assumes that whenever two or more branches are connected in series or in parallel, they are interpreted as only one branch. Hence, before one applies the test, it is convenient to replace all branches which are in series or in parallel by a single branch.

4-8-3 ALGORITHM FOR DRAWING THE DUAL OF A PLANAR NETWORK

Given a planar network N, we can now draw its dual by the following procedure:

1. Check first that all element-characteristic curves are specified with a set of reference directions and polarities consistent with the assumption that the positive terminal of each element is at the tail end of the current reference arrow. If not, redefine the references and the characteristic curves accordingly.
2. Draw the graph of N. Place a node inside each mesh in N and one node outside of N. Corresponding to each branch b_j in N which is common to meshes α and β, draw a branch b_j' from the node inside mesh α to the node inside mesh β. The reference direction for b_j' is then directed toward the node inside mesh α, if the reference direction of branch b_j in mesh α is in the clockwise direction (with respect to the node inside mesh α). Otherwise, the reference direction of b_j' is directed away from the node inside mesh α.
3. The branches b_j' in N' are then replaced by the corresponding network elements dual to branch b_j. For convenience, Table 4-6 tabulates the dual quantities useful for drawing dual networks.

EXAMPLE

Consider the nonlinear bridge network shown in Fig. 4-23*a*. The elements consist of an ideal diode (element 1), a nonlinear resistor (element 2) whose v_2-i_2 curve is shown in Fig. 4-23*b*, a current source $i(t) = 5e^{-t}$ (element 3), a nonlinear capacitor (element 4) whose v_4-q_4 curve is shown in Fig. 4-23*c*, a 5-Ω linear resistor (element 5), a nonlinear inductor (element 6) characterized by $\varphi_6 = 1 + i_6 - 3i_6^2 + 5i_6^3$, a 3-F linear capacitor (element 7), a 2-H linear inductor (element 8), a tunnel diode (element 9) whose v_9-i_9 curve is given in Table 1-1, and a voltage source $v(t) = 2 \cos t$. The units of all voltages, currents, charges, and flux linkages are assumed to be volts, amperes, coulombs, and webers. We draw the dual following the three steps outlined in the above algorithm:

1. A check of the reference directions and polarities of the elements in Fig. 4-23*a* shows that the references for elements 2, 3, 6, and 10 are not consistent with convention. Hence we must

TABLE 4-6 *Common dual quantities.*

Network N	Network N'
Current i_j	Voltage v'_j
Voltage v_j	Current i'_j
Flux linkage φ_j	Charge q'_j
Charge q_j	Flux linkage φ'_j
Nonlinear resistor (characterized by a curve Γ in the v-i plane)	Nonlinear resistor (characterized by the same curve Γ in the i'-v' plane, i.e., with v and i axes interchanged)
Linear resistor with a resistance of R Ω	Linear resistor with a conductance of R mhos or $1/R$ Ω
Nonlinear inductor (characterized by a curve Γ in the i-φ plane)	Nonlinear capacitor (characterized by a curve Γ in the v'-q' plane)
Linear inductor with an inductance of K H	Linear capacitor with a capacitance of K F
Nonlinear capacitor (characterized by a curve Γ in the v-q plane)	Nonlinear inductor (characterized by a curve Γ in the i'-φ' plane)
Linear capacitor with a "capacitance" of K F	Linear inductor with an "inductance" of K H
Voltage source, $v_j = f(t)$	Current source $i'_j = f(t)$
Current source $i_j = g(t)$	Voltage source $v'_j = g(t)$
Short circuit	Open circuit
Open circuit	Short circuit
Parallel branches	Series branches
Series branches	Parallel branches
Link	Tree branch
Tree branch	Link
Fundamental loop	Fundamental cut set
Fundamental cut set	Fundamental loop
Ideal diode	Ideal diode with its two terminals interchanged

redefine the references of these elements as shown in Fig. 4-24*a*. The v-i curve of element 2 must be changed accordingly, as shown in Fig. 4-24*b*. The i-φ curve of element 6 must also be changed to $\varphi_6 = 1 - i_6 - 3i_6^2 - 5i_6^3$, since i_6 has become $-i_6$.

The voltage source and the current source remain unchanged, since the terminal voltage of a voltage source is independent of its terminal current, and the terminal current of a current source is independent of its terminal voltage.

2. The graph of the modified network in Fig. 4-24*a* is drawn in Fig. 4-24*c*, together with its dual graph.
3. The completed dual graph is redrawn as in Fig. 4-24*d*. It remains to replace each branch by its corresponding dual elements. Therefore, element 1 is an ideal diode with its terminals interchanged; element 2 is a nonlinear resistor with its v_2'-i_2' curve shown in Fig. 4-24*f*; element 3 is a voltage source with terminal voltage $v_3'(t) = 5e^{-t}$; element 4 is a nonlinear inductor with its i_4'-φ_4' curve shown in Fig. 4-24*g*; element 5 is a 1/5-Ω linear resistor; element 6 is a nonlinear capacitor characterized by $q_6' = 1 - v_6' - 3v_6'^2 - 5v_6'^3$; element 7 is a 3-H linear inductor; element 8 is a 2-F linear capacitor; element 9 is a nonlinear resistor with its v_9'-i_9' curve shown in Fig. 4-24*h*; and element 10 is a current source with terminal current $i_{10}(t) = 2 \cos t$.

Observe that the dual of a voltage-controlled *v-i* curve is current-controlled, and vice versa. In particular, the dual *v-i* curve can be obtained by reflecting the original curve with respect to the 45° line through the origin. In practice, this operation can be simulated by a 45° reflector (more commonly known as a gyrator).

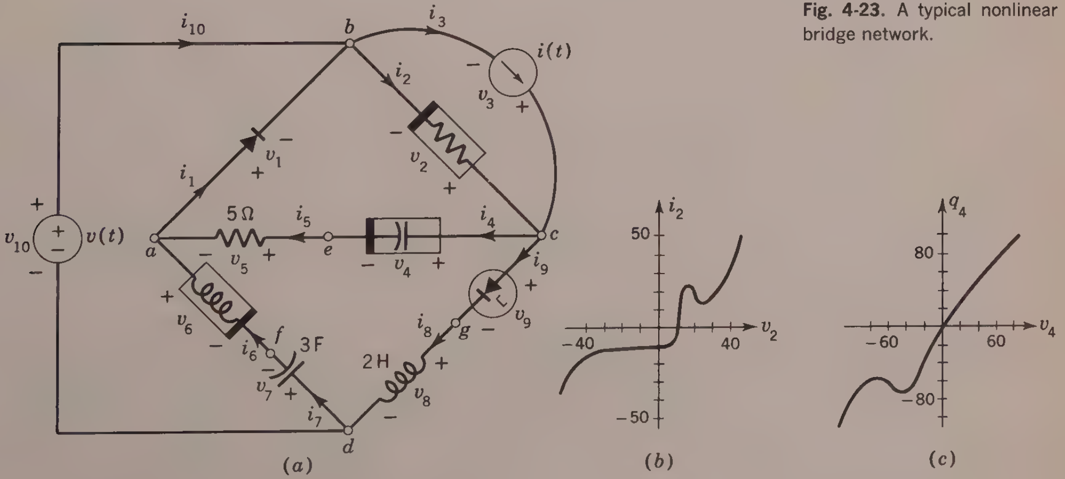


Fig. 4-23. A typical nonlinear bridge network.

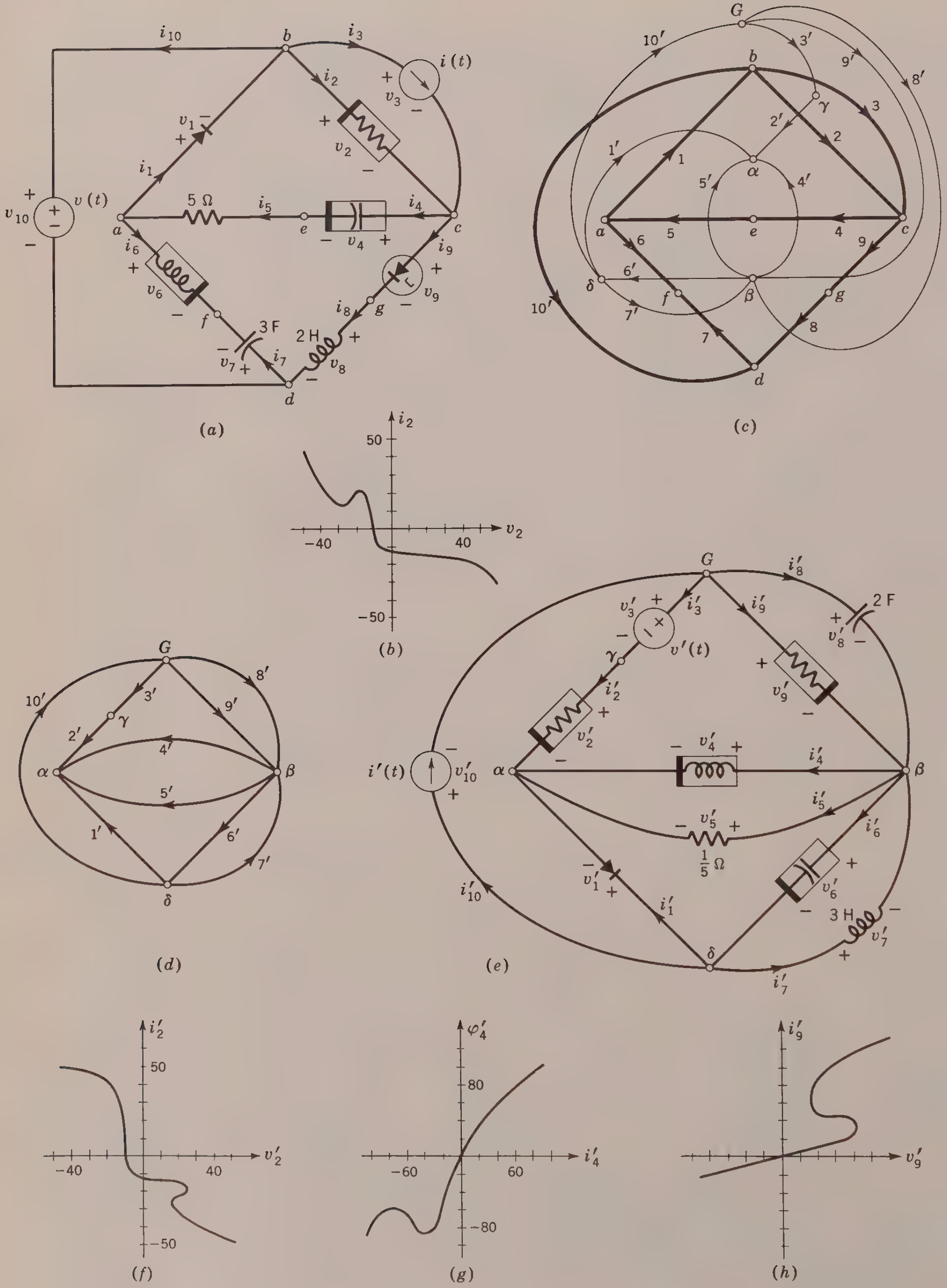


Fig. 4-24. The step-by-step procedure for constructing the dual of the nonlinear bridge network of Fig. 4-23.

Similarly, the type 1 *C-L* mutator presented in Sec. 3-8-3 (also known as a gyrator) can be used to simulate the dual of a nonlinear capacitor and a nonlinear inductor. In other words, the dual of any nonlinear network can be synthesized with the help of a gyrator.

Exercise 1: (*a*) Construct the dual of the network shown in Fig. 4-17*a* by the procedure described in this section. (*b*) Repeat (*a*) for the network shown in Fig. 4-17*b*. (*c*) Show that a planar network can have only one dual.

Exercise 2: Identify the Kuratowski subgraph for the nonplanar network shown in Fig. 4-5.

Exercise 3: Prove that a gyrator can be used to obtain the dual of any nonlinear resistor, inductor, or capacitor.

4-9 SUMMARY

Classification of networks

1. Resistive networks do not contain capacitors or inductors.
   a. Resistive linear networks contain only linear resistors and sources.
   b. Resistive nonlinear networks contain at least one nonlinear resistor.
2. Dynamic networks contain at least one capacitor or inductor.
   a. Dynamic linear networks contain only linear elements and sources.
   b. Dynamic nonlinear networks contain at least one nonlinear element.

Equations of motion

1. Equations from the laws of elements representing the characteristic curves of the elements are always independent.
2. Equations from the laws of interconnection representing equations from KVL and KCL may not be independent.

Network topology

Network topology provides the systematic techniques for obtaining independent equations from the laws of interconnection. Some basic terminologies are as follows:

1. Tree: a set of branches, called tree branches, connecting all nodes but not forming closed loops.

2. Cotree: the set of all branches, called links, not belonging to a tree.
3. Cut set: a set of branches which, if cut, would separate the network into two parts.
4. Fundamental loop: given a tree T, a loop formed by a link of T and one or more tree branches is called a fundamental loop with respect to T.
5. Fundamental cut set: given a tree T, a cut set formed by a tree branch of T and one or more links is called a fundamental cut set with respect to T.

Independent KVL equation criteria The maximum number of independent KVL equations is equal to $b - (n - 1)$, where b is the total number of branches and n is the total number of nodes. These equations can always be written around the $b - (n - 1)$ fundamental loops with respect to a tree T. (These criteria are valid only for networks containing two-terminal elements.)

Generalized KCL The algebraic sum of all branch currents belonging to a cut set is zero.

Independent KCL equation criteria The maximum number of independent KCL equations is equal to $n - 1$ where n is the total number of nodes. These equations can always be written across the fundamental cut sets with respect to a tree T. (These criteria are valid only for networks containing two-terminal elements.)

Ground rule for networks with multiterminal elements If each KVL or KCL equation contains at least one branch variable which did not appear in the preceding equations, then the equations are independent.

Nature of equations of motion

1. Resistive Nonlinear Networks: a system of nonlinear functional equations.
2. Dynamic Nonlinear Networks: a system of nonlinear functional-differential equations. These equations must be recast into the normal form before they are amenable to existing solution techniques. The dependent variables in the normal form are called state variables.

Practical methods for solving nonlinear functional equations

1. Numerical methods
2. Graphical methods (Chaps. 6 and 7)
3. Piecewise-linear methods (Chap. 12)

Some numerical techniques

1. Fixed-point Algorithm: The solution of $X = F(X)$ is equal to the fixed point of $F(X)$. It can be found by iterating $X_{i+1} = F(X_i)$, $i = 0, 1, \ldots, n$.
2. Newton-Raphson Method: The solution of $G(X) = 0$ is equal to the fixed point of $F(X) = X - [G'(X)]^{-1}G(X)$. It can be found by iterating $X_{i+1} = X_i - [G'(X_i)]^{-1}G(X_i)$.

Practical methods for solving nonlinear differential equations

1. Numerical methods
2. Graphical methods (Chaps. 18 and 19)
3. Piecewise-linear methods (Chaps. 14 to 17)
4. Analog simulation

Euler algorithm, a simple numerical technique

Given $dx/dt = f(x,t)$ and $x(t_0) = x_0$, corresponding to a step size h and a total of n iterations, the approximate solution is given by

$$x_{i+1} = x_i + hf(x_i,t_i)$$

$$t_{i+1} = t_i + h$$

$$i = 0, 1, 2, \ldots, n - 1$$

Principles of duality

1. Only planar networks have a dual. (See construction procedure in the text.)
2. Any dual element can be realized with the help of gyrators.

PROBLEMS

4-1 Consider the network N shown in Fig. P4-1.
   (a) Draw the network graph of N. Label all branches and nodes.
   (b) Exhibit a tree and list the branches of the cotree corresponding to your tree.

(c) List the fundamental loops corresponding to the tree chosen in (*b*).
(d) List the fundamental cut sets corresponding to the tree chosen in (*b*).
(e) Find the number of independent KCL and KVL equations for the network *N*.

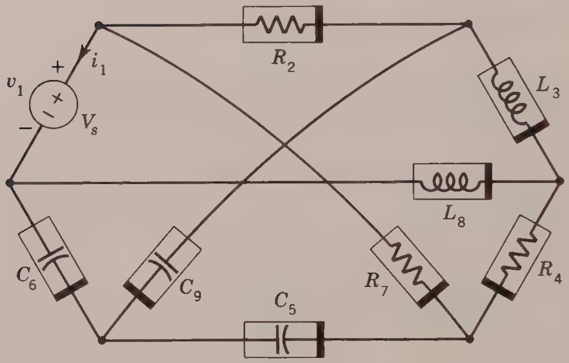


Fig. P4-1.

4-2 Prove the independent KVL mesh equation criteria. HINT: To prove the first part, use the principle of mathematical induction. Start with $b - (n - 1) = 1$ and then $b - (n - 1) = k$. To prove the second part, show that if the mesh equations are not independent, there exists at least one loop equation independent of the mesh equations. Then show that any loop equation can be expressed as a linear combination of mesh equations, thereby establishing a contradiction.

4-3 Prove the independent KCL equation criteria. HINT: Apply duality concepts to the proof of the independent KVL equation criteria.

4-4 If a network is planar, it is usually easier to write the KVL equations around the meshes of the network. Apply this method to derive the equations of motion of the network shown in Fig. 4-7*a*.

4-5 Write the equations of motion of the circuit shown in Fig. P4-5. The nonlinear resistors R_2 and R_6 are junction diodes characterized

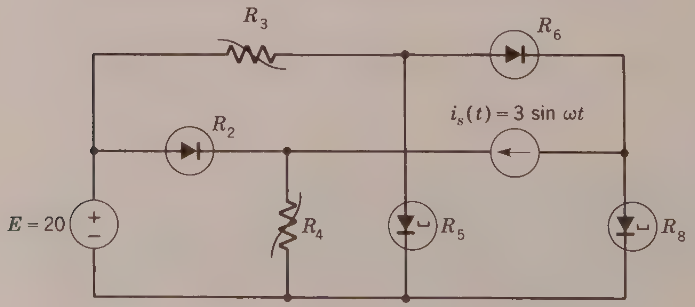


Fig. P4-5.

by a v-i curve $i = I_o(e^{kv} - 1)$. The nonlinear resistors R_3 and R_4 are varistors characterized by a v-i curve $v = 100i^{1/3}$. The nonlinear resistors R_5 and R_8 are tunnel diodes characterized by a v-i curve $i = 10.1v - 26.5v^2 + 17.4v^3$. Eliminate as many variables as possible.

4-6 Find the equations of motion of the nonlinear network shown in Fig. P4-6. The two-terminal nonlinear resistor is characterized by $i_2 = 2v_2^3$. The three-terminal nonlinear resistor is characterized by $i_5 = f_1(v_5,i_6)$ and $v_6 = f_2(v_5,i_6)$. Eliminate as many variables as possible.

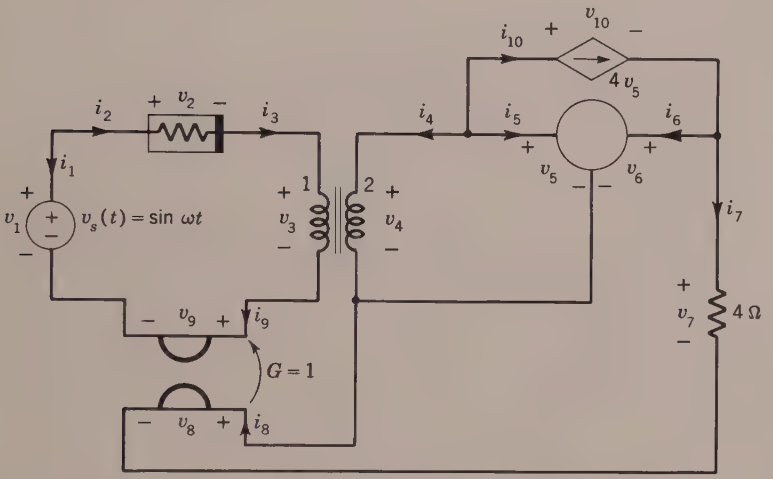


Fig. P4-6.

4-7 Consider the nonlinear equation $g(x) = x - (1/4)x^2 = 0$ and its associated function $f(x) = x - k(x)g(x)$.

(a) With $k(x) = 1$, find the fixed points of $f(x)$ analytically. Do they agree with the solutions of $g(x) = 0$? If not, explain why.

(b) Repeat (*a*) with $k(x) = x^2 - 4$.

(c) Assuming $k(x) = 1$, carry out the fixed-point iteration four times with the initial guess $x_0 = 2$. Does the iteration converge?

(d) Repeat (*c*) with $x_0 = -3$.

(e) Repeat (*c*) with $x_0 = 5$.

(f) Can you find an appropriate $k(x)$ and some appropriate initial guesses so that all fixed points can be obtained by iteration?

4-8 Find the solution of the equation $g(x) = x - (1/4)x^2 = 0$ by the Newton-Raphson method. Carry out the iteration four times with each of the initial guesses $x_0 = 2$, -3, and 5. Compare its rate of convergence with the results in Prob. 4-7.

4-9 It can be shown that if the magnitude of the slope of $f(x) \equiv x - kg(x)$ is less than a constant $n < 1$, the fixed-point algorithm will always converge. If $g(x) = 0$ is defined over the interval $a \leq x \leq b$, where $g(a) < 0$ and $g(b) > 0$, find the conditions under which a constant k may be found to guarantee $|f'(x)| < 1$ over the interval $a \leq x \leq b$. Give a numerical example to illustrate how to choose the value k.

4-10 The equations of motion of an instrumentation circuit have been written and simplified to the following two equations:

$$v_1 - v_1^2 + v_2^2 = 0$$

$$v_1 - 4v_2 = 0$$

(a) Find the solution analytically.
(b) Find the solution graphically.
(c) Recast these equations into the fixed-point form for two equations and carry out the iteration with the initial guess $(v_1, v_2) = (0.25, -0.25)$.

4-11 Figure P4-11 shows a series nonlinear RLC network. R is described by $i_1 = 2 \tanh v_1$, L by $\varphi_2 = 3i_2^3$, C by $q_3 = \ln v_3$, and the voltage source by $v_s(t) = 4e^{-t} \sin t$ volts. Find the equations of motion in the normal form in terms of the following state variables:
(a) i_2 and v_3
(b) i_2 and q_3
(c) φ_2 and v_3
(d) φ_2 and q_3

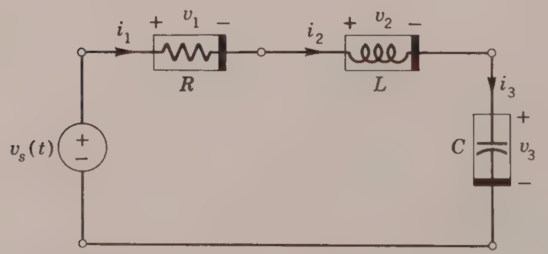


Fig. P4-11.

4-12 Solve Prob. 4-11 for the case when the inductor characteristic is given by (*a*) $\varphi_2 = i_2^3 + i_2^2$ and (*b*) $i_2 = \varphi_2^3 + \varphi_2^2$. NOTE: The normal form may not exist for some choice of state variables.

4-13 Consider the dynamic network shown in Fig. P4-13, where the three-terminal resistor is characterized by

$$v_1 = 2 + 3v_2 + v_2 i_1^3 - 2i_1^5$$

$$i_2 = 4 - i_1^2 v_2 - 2i_1 v_2^3 + 3v_2^5$$

(a) Write the equations of motion if the inductor and capacitor are characterized by

$$\varphi = 2 - 2i^3 + 5i^5$$
$$q = 1 + 2v - 3v^3 + v^5$$

(b) Choose a set of state variables in (*a*) and express the equations of motion in the normal form.

(c) Repeat (*a*) for the case where the inductor and capacitor are characterized by

$$i = 1 + \varphi + \varphi^2$$
$$v = q - q^3$$

(d) Choose a set of state variables in (*c*) and express the equations of motion in the normal form.

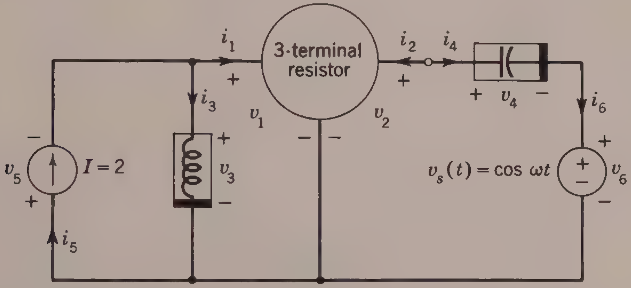


Fig. P4-13.

4-14 Consider the differential equation $dx/dt = f(x,t)$ and the iteration formula

$$x_{i+1} = x_i + h\left\{\frac{f(x_i,t_i) + f[x_i + hf(x_i,t_i),t_{i+1}]}{2}\right\}$$

$$t_{i+1} = t_i + h$$
$$i = 0, 1, 2, \ldots, n-1$$

(a) The above formula is called the *Heun algorithm.* Give a geometrical interpretation of this algorithm and explain why it represents a more accurate method than the Euler algorithm.

(b) Find the solution of Eq. (4-63) by the Heun algorithm with the same initial condition and step size as in the text. Compare the accuracy with that obtained by the Euler algorithm.

(c) Extend this algorithm to a system of *n* differential equations.

4-15 Another numerical technique for solving the differential equation $dx/dt = f(x,t)$ is the *Taylor algorithm* defined by

$$x_{i+1} = x_i + hf(x_i,t_i) + \frac{h^2}{2!}f'(x_i,t_i) + \frac{h^3}{3!}f''(x_i,t_i) + \cdots + \frac{h^m}{m!}f^{(m-1)}(x_i,t_i)$$

$$t_{i+1} = t_i + h$$
$$i = 0, 1, 2, \ldots, n-1$$

(a) Give a geometrical interpretation of this method and explain why it is more accurate than the Euler algorithm.
(b) Find the solution of Eq. (4-63) by the Taylor algorithm with the same initial condition and step size as in the text.
(c) What is the disadvantage of this method over the Euler algorithm?

4-16 One of the most commonly used algorithms for solving the differential equation $dx/dt = f(x,t)$ without the necessity of evaluating high-order derivatives is the *Runge-Kutta algorithm* defined by

$$x_{i+1} = x_i + hg(x_i,t_i)$$

where

$$g(x_i,t_i) \equiv \frac{1}{6}(k_1 + 2k_2 + 2k_3 + k_4)$$

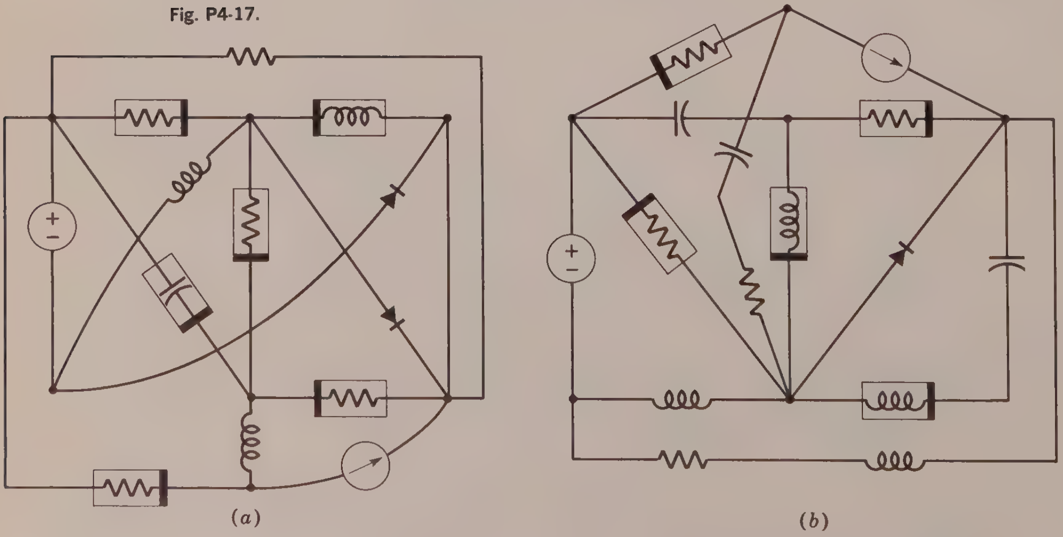
$$k_1 = f(x_i,t_i) \qquad k_2 = f\left(x_i + \frac{h}{2}k_1, t_i + \frac{h}{2}\right)$$

$$k_3 = f\left(x_i + \frac{h}{2}k_2, t_i + \frac{h}{2}\right) \qquad k_4 = f(x_i + hk_3, t_i + h)$$

(a) Give a geometrical interpretation of this algorithm.
(b) Find the solution of Eq. (4-63) by the Runge-Kutta algorithm with the same initial condition and step size as in the text. Compare its accuracy with that of the Euler algorithm.

4-17 Test whether the networks shown in Fig. P4-17*a* and *b* are planar or nonplanar. Justify your answers.

Fig. P4-17.



4-18 (a) Find the dual of the nonlinear bridge-T network shown in Fig. P4-18*a*. The *v-i* curves of the resistors are given in

Table 1-1 of Chap. 1. The inductor i-φ curve is represented by $i = \varphi^3$, and the capacitor v-q curve is represented by $v = \tanh q$.

(b) Repeat (*a*) for the nonlinear twin-T network shown in Fig. P4-18*b*.

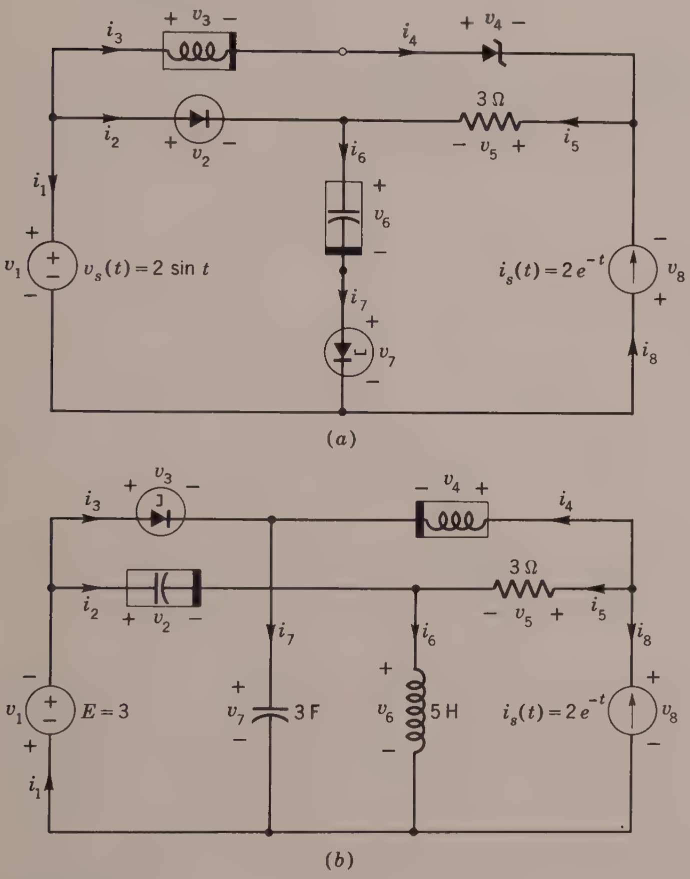


Fig. P4-18.

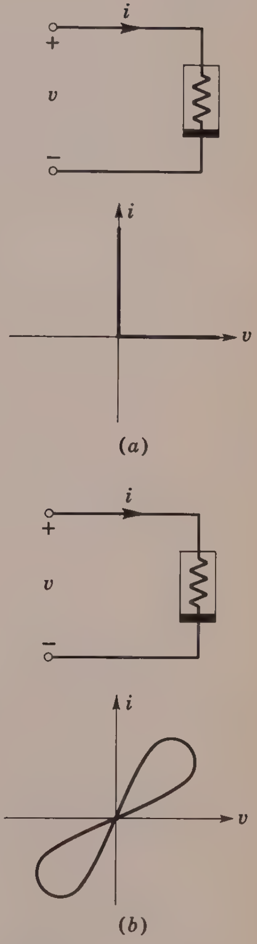


Fig. P4-19.

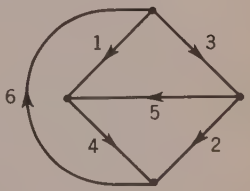
4-19 An element whose dual is itself is called a self-dual element. Show that the elements of Fig. P4-19*a* and *b* are self-dual elements. What property must a v-i curve possess in order for it to represent a self-dual element?

4-20 If a network graph and its dual have the same topology, the graph is said to be a self-dual graph.

(a) Verify that the graph of Fig. P4-20 is a self-dual graph.

(b) Find two more self-dual graphs.

Fig. P4-20.



part 2

RESISTIVE NONLINEAR NETWORKS

5 THREE BASIC CONCEPTS OF RESISTIVE NONLINEAR NETWORKS

5-1 THE OPERATING-POINT CONCEPT

We have shown in Sec. 4-4 that the equations of motion of any resistive nonlinear network N can always be written and reduced into the form of a system of independent, functional (algebraic or transcendental) equations of the form

$$\begin{aligned} &f_1(v_1,v_2,\ldots,v_m,i_1,i_2,\ldots,i_n,t) = 0 \\ &f_2(v_1,v_2,\ldots,v_m,i_1,i_2,\ldots,i_n,t) = 0 \\ &\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ &f_{m+n}(v_1,v_2,\ldots,v_m,i_1,i_2,\ldots,i_n,t) = 0 \end{aligned} \qquad (5\text{-}1)$$

The parameter t will be present whenever one or more of the sources in N is time-dependent. The solution of Eq. (5-1) then consists of a set of time functions

$$v_j = E_j(t) \qquad j = 1,\ldots,m$$

and

$$i_j = I_j(t) \qquad j = 1,\ldots,n$$

such that, when these are substituted into Eq. (5-1), the left side of each equation will be identically zero for all time t, that is,

$$\begin{aligned} &f_1(E_1(t),E_2(t),\ldots,E_m(t),I_1(t),I_2(t),\ldots,I_n(t),t) \equiv 0 \\ &f_2(E_1(t),E_2(t),\ldots,E_m(t),I_1(t),I_2(t),\ldots,I_n(t),t) \equiv 0 \\ &\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ &f_{m+n}(E_1(t),E_2(t),\ldots,E_m(t),I_1(t),I_2(t),\ldots,I_n(t),t) \equiv 0 \end{aligned} \qquad (5\text{-}2)$$

Since Eq. (5-1) is assumed to be already in its reduced form, i.e., all possible elimination and substitution have been made, we must now solve for the unknowns by the nonanalytical techniques described in Sec. 4-5, namely, numerical, graphical, and piecewise-linear. If Eq. (5-1) is given in analytic form, and if there is a unique solution, then the numerical method is usually superior, provided the iteration procedure converges.[1] In practice, however, Eq. (5-1) is really just a symbolic notation because each equation would generally contain functions specified by curves, not equations. In such cases, the graphical and the piecewise-linear techniques are far superior.[2]

An examination of Eq. (5-1) shows that the presence of the parameter t greatly increases the difficulty of the problem. Let us therefore consider the simpler problem of finding the solution of Eq. (5-1) at a specific time $t = t_0$ rather than for all time t. By substituting $t = t_0$ in Eq. (5-1), the parameter t can be eliminated. This is equivalent to replacing all time-dependent sources in N by dc sources whose values are equal to the values of the time-dependent sources at t_0. In other words, N becomes a dc-resistive network,[3] and Eq. (5-1) becomes

$$\begin{array}{l} f_1(v_1,v_2,\ldots,v_m,i_1,i_2,\ldots,i_n) = 0 \\ f_2(v_1,v_2,\ldots,v_m,i_1,i_2,\ldots,i_n) = 0 \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ f_{m+n}(v_1,v_2,\ldots,v_m,i_1,i_2,\ldots,i_n = 0 \end{array} \tag{5-3}$$

If we can find the solution of Eq. (5-3), then we can find the solution of Eq. (5-1) for all time t by simply solving the problem at different instants of time. In other words, if we can solve a dc-resistive network, then we can also solve the corresponding ac-resistive network. This observation shows that the fundamental problem in nonlinear network theory is the solution of dc-resistive networks, and therefore only this class of resistive networks will be considered in the rest of this chapter.

The solution of Eq. (5-3) consists of a set of numbers $(E_1,E_2,\ldots,E_m,I_1,I_2,\ldots,I_n)$ such that

$$\begin{array}{l} f_1(E_1,E_2,\ldots,E_m,I_1,I_2,\ldots,I_n) = 0 \\ f_2(E_1,E_2,\ldots,E_m,I_1,I_2,\ldots,I_n) = 0 \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ f_{m+n}(E_1,E_2,\ldots,E_m,I_1,I_2,\ldots,I_n) = 0 \end{array} \tag{5-4}$$

---

[1] Unfortunately, no single numerical method which can guarantee convergence is presently known.

[2] These methods can be easily implemented by a computer.

[3] A network containing only dc sources is called a dc network in this book. In particular, we shall distinguish between a *dc-resistive network* and a *dc-dynamic network*. By tradition, a source is said to be an ac source (derived from the words alternating current) if it varies with time. Accordingly, a network containing one or more ac sources is called an *ac-resistive network* or an *ac-dynamic network*.

Conceptually, it is useful to interpret Eq. (5-3) and its solution from a geometric point of view. First, let us consider Eq. (5-3) in its simplest form, namely, two equations in two unknowns v_1 and i_2; thus

$$\begin{aligned} f_1(v_1,i_2) &= 0 \\ f_2(v_1,i_2) &= 0 \end{aligned} \tag{5-5}$$

These equations represent two curves in the two-dimensional v_1-i_2 plane. The intersection (E_1,I_2) of the two curves is a solution of Eq. (5-5) because it satisfies both equations. Hence we can say that a solution of Eq. (5-5) is a point (E_1,I_2) in a two-dimensional space. Next, consider Eq. (5-3) in three variables; we have

$$\begin{aligned} f_1(v_1,v_2,i_1) &= 0 \\ f_2(v_1,v_2,i_1) &= 0 \\ f_3(v_1,v_2,i_1) &= 0 \end{aligned} \tag{5-6}$$

Each of these equations represents a surface in the three-dimensional v_1-v_2-i_1 space. An intersection (E_1,E_2,I_1) of the three surfaces is a solution of Eq. (5-6) because it satisfies all three equations. Hence, we can say that a solution of Eq. (5-6) is a point (E_1,E_2,I_1) in a three-dimensional space. By a logical generalization of these geometrical interpretations, we can interpret each equation in Eq. (5-3) as a surface in the $(m + n)$-dimensional v_1-v_2 $\cdots$ -v_m-i_1-i_2 $\cdots$ -i_n space, and each point of intersection $(E_1,E_2, \ldots ,E_m; I_1,I_2, \ldots ,I_n)$ of these $(m + n)$ surfaces is clearly a solution of Eq. (5-3). Hence, we can say that a solution of Eq. (5-3) is a point $(E_1,E_2, \ldots ,E_m; I_1,I_2, \ldots ,I_n)$ in an $(m + n)$-dimensional space. We shall call such a point of intersection an *operating point* of the network N. The operating point of a network N is simply a set of numbers (representing a solution of N) which we interpret as the coordinates of a point. Since "surfaces" in general may intersect at only one point, at finitely many points, at infinitely many points, or sometimes not at all, it is clear that a network N may have one operating point, finitely many operating points, infinitely many operating points, or sometimes no operating points at all. This is equivalent to saying that a network N may, in general, have one solution, finitely many solutions, infinitely many solutions, or sometimes no solution at all. The physical interpretation of networks containing more than one operating point is discussed in Sec. 5-5. We shall refer to the problem of finding the operating points of a dc-resistive network as the *operating-point problem.*

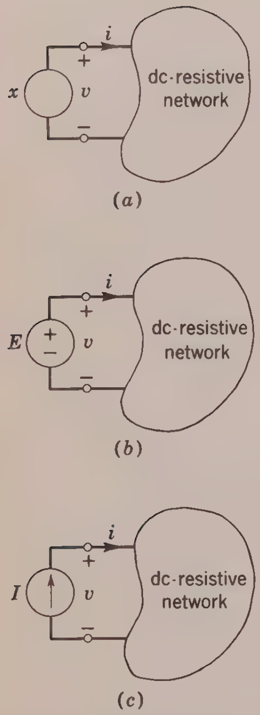


Fig. 5-1. A dc-resistive network with an external ac source whose terminal voltage or terminal current is being varied.

5-2 CONCEPTS OF DRIVING-POINT AND TRANSFER-CHARACTERISTIC PLOTS

We shall now introduce two concepts of fundamental importance in nonlinear network theory. Most of the material in the remaining chapters of this book will depend heavily on the concepts introduced in this section. Consider an arbitrary dc-resistive nonlinear network N which may contain n-terminal resistors, controlled sources, ideal transformers, gyrators, and dc sources. An operating point P of N is then a set of numbers which we have interpreted to be the coordinates of P, that is,

$$P:(E_1,E_2,\ldots,E_m,I_1,I_2,\ldots,I_n) \tag{5-7}$$

Suppose we now change the values of some of the dc sources in N. Clearly, the network will have a new operating point P' with a new set of coordinates

$$P':(E_1',E_2',\ldots,E_m',I_1',I_2',\ldots,I_n') \tag{5-8}$$

More specifically, let us change the value of only one of the dc sources in N. In order to emphasize that only one source is being altered, we shall draw the network in the form shown in Fig. 5-1*a*, where we have denoted the source that is being altered by x. Observe that x may be a voltage source or a current source, and thus Fig. 5-1*a* may become either Fig. 5-1*b* or Fig. 5-1*c*. Let us now vary the value of the source x from $-\infty$ to $+\infty$. Corresponding to each value of x, we shall obtain one or more operating points[1] in the $(m+n)$-dimensional space. If $(m+n)=3$, we can plot each point P corresponding to each value of x and denote it by $P(x)$, as shown in Fig. 5-2, where the operating points have been connected into a space curve, henceforth referred to as the *locus of operating points.*

It is not difficult to imagine that when $(m+n)>3$, the locus of operating points will still be a curve in the $(m+n)$-dimensional space. Although we cannot plot the locus in an $(m+n)$-dimensional space, we can nevertheless specify the locus of operating points by several means. The simplest method is to tabulate the coordinates of each point P corresponding to each parameter x in the form of a table such as Table 5-1. A disadvantage of the table,

[1] We have assumed here that N has at least one operating point. The special case where N has no solution is of no practical interest.

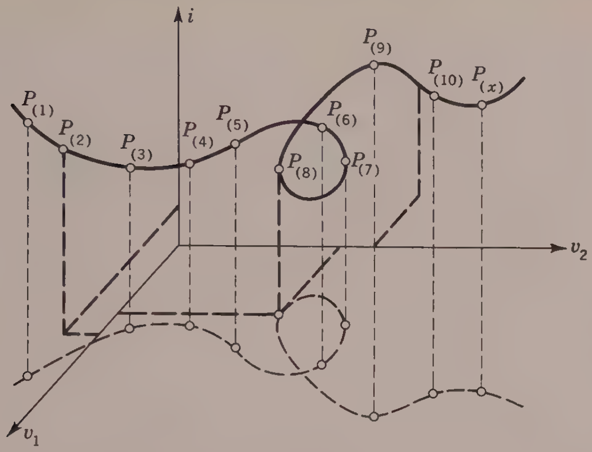


Fig. 5-2. The locus of the operating points in a three-dimensional space corresponding to different terminal values of the driving source x can be plotted as a space curve with x as a parameter.

however, is that only a finite number of operating points can be tabulated. Hence, in Table 5-1 we can find the operating points corresponding to $x = 1, 2, 3$, etc., but we cannot find the operating point corresponding to $x = 2.3, 3.7$, etc. It is difficult to interpolate the intermediate values from the table, because, as will be seen later, the relationships between the coordinates of P and the values of x are usually nonlinear.

A more accurate method of specifying the locus of operating points consists of plotting *each coordinate* of the operating point as a function of x in the form of a curve. Each curve can be plotted from the data in the appropriate column of Table 5-1. For example, the data in the column corresponding to the coordinate v_j in Table 5-1 can be plotted as shown in Fig. 5-3*a*. These points can then be used to draw a smooth curve.[1]

Similarly, the data in the column corresponding to i_j can be plotted as shown in Fig. 5-3*b*. Notice that in order to completely specify the locus of the operating points, we would need as many curves as there are columns in Table 5-1.

Now observe that each voltage coordinate v_j of the operating point P corresponds to the voltage across some pair of nodes in N; similarly, each current coordinate i_j of the operating point P corresponds to the current in some wire. For example, v_j might represent the voltage between nodes *a-b*, and i_j might represent the current in wire *c-d* of the network shown in Fig. 5-4*a*. In order to emphasize that the curve v_j versus x represents the voltage solution across nodes *a-b* as a function of x, and the curve i_j versus x represents the current solution in the wire *c-d* as a function of x, let us redraw Fig. 5-4*a* into the more explicit form shown in Fig.

---

[1] It is possible for the curve to jump abruptly, as for a discontinuity. This, however, will not occur in physical networks.

TABLE 5-1 *The locus of operating points of a typical resistive network.*

Value of x	Coordinates of operating point P												
	v_1	v_2	v_3	...	v_j	...	v_m	i_1	i_2	...	i_j	...	i_n
.	.	.	.	...	.	...	.	.	.	...	.	...	.
.	.	.	.	...	.	...	.	.	.	...	.	...	.
.	.	.	.	...	.	...	.	.	.	...	.	...	.
−4	3.2	−3.2	0	...	−2.9	...	−2.3	0	−3.4	...	−3	...	−4.6
−3	2.1	−2.3	−1.9	...	−1.8	...	−0.5	1.6	−2	...	−2.2	...	−2.8
−2	1.5	−1.7	−1.1	...	−2.2	...	0.5	2.5	−2.2	...	1.1, 0.5, −1.7	...	−1.6
−1	1.1	−1.1	0	...	−1.2	...	1.0	1.5	−3.0	...	1.5, −0.6, −1.5	...	−2.4, 0.2, 1.9
0	0.8	−0.8	1.0	...	0	...	1.2	0	−2.7	...	2.0	...	−3.2, −0.4, 2.5
1	0	0	1.2	...	1	...	1.8	−1.5	−1.5	...	2.2	...	−3.8, −0.8, 3.0
2	−2	1.3	0	...	1.2	...	2.6	−1.5	0	...	2.8	...	−3.8, −1.8, 3.5
3	−3.5	2.8	−3.5	...	1.4	...	4.0	0	1.7	...	3.4	...	4.0
4	−4.0	3.9	−4.2	...	2.1	...	4.5	1.6	4.0	...	4.3	...	4.5
.	.	.	.	...	.	...	.	.	.	...	.	...	.
.	.	.	.	...	.	...	.	.	.	...	.	...	.
.	.	.	.	...	.	...	.	.	.	...	.	...	.

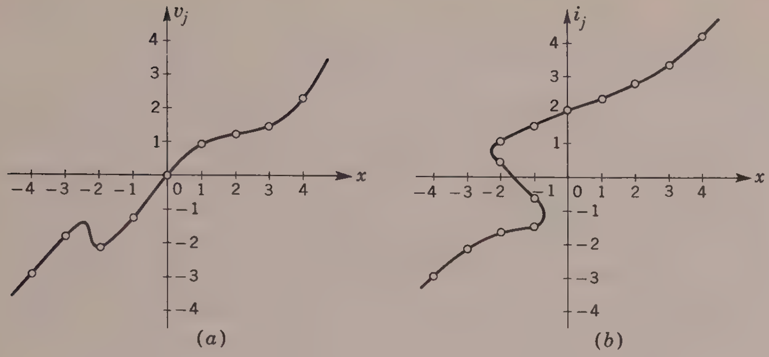
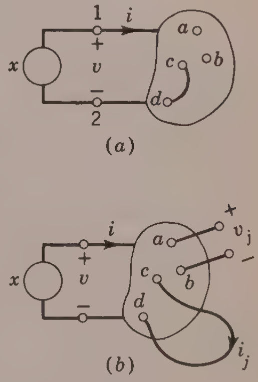


Fig. 5-3. The relationship between each voltage coordinate v_j or current coordinate i_j and the value of the driving source x can be plotted in the form of a plane curve.

5-4*b*. We cannot overemphasize here that the two wires connected to nodes *a-b* as shown in Fig. 5-4*b* are not part of the original network, and hence must be left open-circuited. These two wires are brought out of the box just to emphasize that if we connect a voltmeter across them and plot the measured voltage as we vary the value x, we would obtain the relationship v_j versus x. Clearly, if we connect an external load to these wires (which results in a current flowing in the load), we change the network. As a result of this, we expect to measure a different voltage v_j for the same value of x. Therefore, in order to obtain a v_j-vs.-x relationship unambiguously, it is necessary that no current be allowed to flow in the wires where the output voltage v_j is to be measured.

Depending upon the location of the pair of nodes *a-b* across which the voltage v_j is to be measured and the location of the wire *c-d* through which the current i_j is to be measured, we shall classify the v_j-vs.-x and the i_j-vs.-x curves into two main classes, the *driving-point characteristic plot* and the *transfer-characteristic plot.*

Fig. 5-4. A resistive network drawn in the form of a black box to emphasize the relationship between v_j versus x and i_j versus x.



5-2-1 THE DRIVING-POINT CHARACTERISTIC PLOT (DP PLOT)

Any curve, v_j versus x or i_j versus x, is said to be a driving-point characteristic plot (henceforth referred to as a DP plot) if v_j and x or i_j and x are measured across the same pair of terminals, called the driving-point terminals.

From this definition, we see that there are two possibilities for a v_j-vs.-x or an i_j-vs.-x curve to be a DP plot:

1. When x is a voltage source and i_j corresponds to the current entering one of the two driving-point terminals as shown in Fig. 5-5*a*

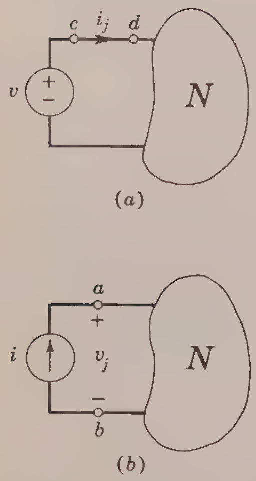


Fig. 5-5. The DP plot pertains to the relationship between the current i_j and the driving voltage v, or the voltage v_j and the driving current i, across the same terminal pair, the driving-point terminals.

2. When x is a current source and v_j corresponds to the voltage across the driving-point terminals shown in Fig. 5-5*b*

In case 1 the DP plot corresponds to the i_j-vs.-v curve, and in case 2 the DP plot corresponds to the v_j-vs.-i curve. Observe that in either case the DP plot is simply the v-i characteristic measured across the driving-point terminals of the resistive network *N*. In other words, the entire dc-resistive network *N* shown in Fig. 5-5*a* or *b* can be interpreted as a two-terminal resistor characterized by a v-i curve, which also happens to be the DP plot of the network.

5-2-2 THE TRANSFER-CHARACTERISTIC PLOT (TC PLOT)

Given a dc-resistive network *N*, there will be many occasions when we would like to "create" a pair of driving-point terminals in *N* and consider the resulting black box as a two-terminal resistor. We refer to the creation of new driving-point terminals as finding an "entry" to the network. We can obtain an entry to a network either by the "soldering-iron" entry or by the "pliers-type" entry. In the "soldering-iron" entry we simply choose two convenient nodes which already exist in *N* and "solder" to these nodes a pair of external terminals. The external terminals can connect to either a voltage source as in Fig. 5-6*a* or a current source as in Fig. 5-6*b*. In the "pliers-type" entry we choose a convenient wire in *N* and cut it with a "pliers." The result is two newly created terminals which can be connected across either a voltage source as shown in Fig. 5-7*a* or a current source as shown in Fig. 5-7*b*.

As an illustration of the creation of driving-point terminals, consider the simple dc-resistive network *N* shown in Fig. 5-8. We can create three different driving-point terminals in the network *N* by soldering-iron entries, as shown in Fig. 5-9*a* to *c*. Similarly, we can create six different driving-point terminals to the network *N* by pliers-type entries, three of which are shown in Fig. 5-10*a* to *c*.

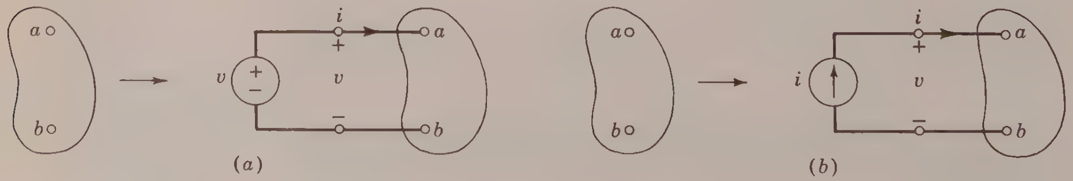


Fig. 5-6. A soldering-iron entry is obtained by soldering a pair of terminals across a pair of existing nodes.

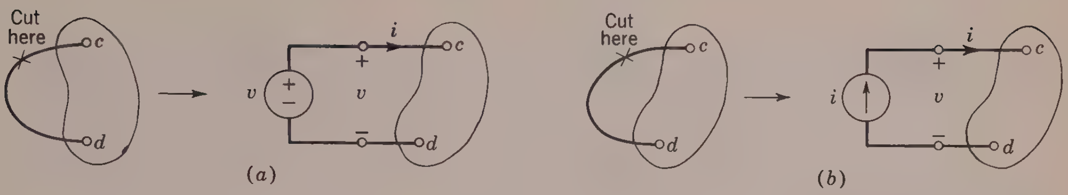


Fig. 5-7. A pliers-type entry is obtained by cutting a wire, thereby creating a pair of terminals.

DEFINITION OF A TRANSFER-CHARACTERISTIC PLOT (TC PLOT)

Any curve v_j versus x or i_j versus x is said to be a transfer-characteristic plot (henceforth referred to as a TC plot) if v_j is measured across one pair of terminals and x is measured across another or if i_j is measured entering one terminal and x is measured across another. We shall refer to the terminal pair corresponding to x as the *driving-point terminals* and to the terminal pair corresponding to v_j or i_j as the *transfer terminals.* With reference to Fig. 5-4 and the fact that x may be either a voltage source or a current source, it is clear that there are four different types of TC plots. If we denote the driving-point terminals by the subscript k and the transfer terminals by the subscript j, then the types of TC plots are as follows:

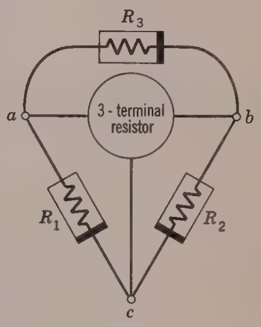


Fig. 5-8. A typical dc-resistive network.

1. v_j-vs.-v_k plot. This corresponds to the case where the driving-point variable is v_k and the transfer variable is v_j, as shown in Fig. 5-11*a*. An example of this type of TC plot, shown in Fig. 5-11*b*, was obtained by plotting the data in column v_1 of Table 5-1 with respect to $x = v_k$.

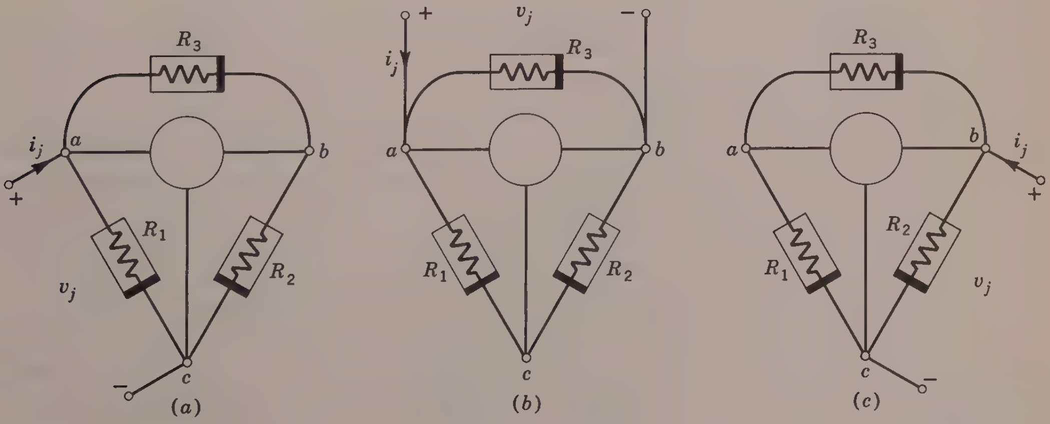


Fig. 5-9. Examples of three pairs of driving-point terminals created by the soldering-iron entry.

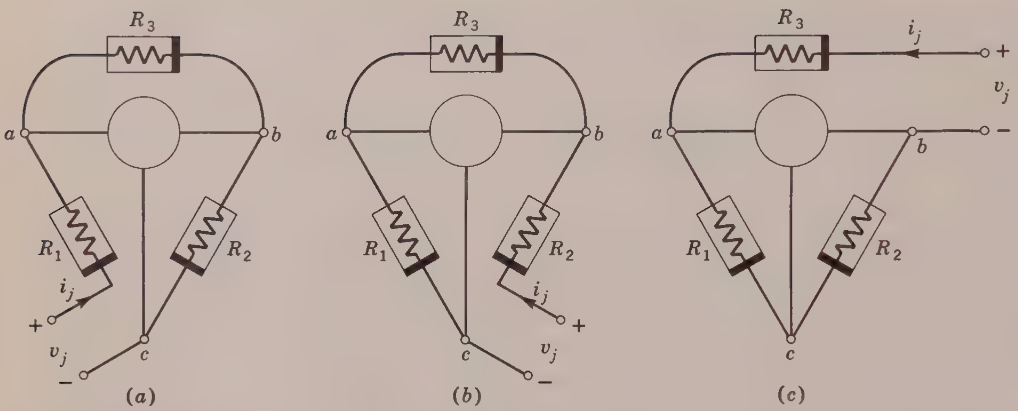


Fig. 5-10. Examples of three of the six possible driving-point terminals created by the pliers-type entry.

2. v_j-vs.-i_k plot. This corresponds to the case where the driving-point variable is i_k and the transfer variable is v_j, as shown in Fig. 5-11*c*. An example of this type of TC plot is shown in Fig. 5-11*d*, which was obtained by plotting the data in column v_2 of Table 5-1 with respect to $x = i_k$.

3. i_j-vs.-v_k plot. This corresponds to the case where the driving-point variable is v_k and the transfer variable is i_j, as shown in Fig. 5-11*e*. An example of this type of TC plot, obtained by plotting the data in column i_2 (that is, with $j = 2$) with respect to $x = v_k$, is shown in Fig. 5-11*f*.

4. i_j-vs.-i_k plot. This corresponds to the case where the driving-point variable is i_k and the transfer variable is i_j, as shown in Fig. 5-11*g*. Figure 5-11*h* is an example of this type of TC plot, obtained by plotting the data in column i_n (that is, with $j = n$) with respect to $x = i_k$.

The definitions of DP and TC plots actually tell us that any v_j-vs.-x or i_j-vs.-x curve which is not a DP plot must be a TC plot. Observe also that in the v_j-vs.-x plots shown in Fig. 5-11*b* and *d*, v_j corresponds to the voltage across the transfer terminals with $i_j = 0$. We emphasize again that any v_j-vs.-x curve describing a transfer characteristic is valid only under the condition that the transfer terminals are left open, thus ensuring that $i_j = 0$. If we connect across the transfer terminals a two-terminal black box often referred to as the *load*, then $i_j \neq 0$, and the original v_j-vs.-x TC plot will be altered. The amount of change or discrepancy from the original plot is often referred to as the *loading effect*. Similarly, the i_j-vs.-x plots shown in Fig. 5-11*f* and *h* are valid only if $v_j = 0$, that is, the transfer terminals are short-circuited.

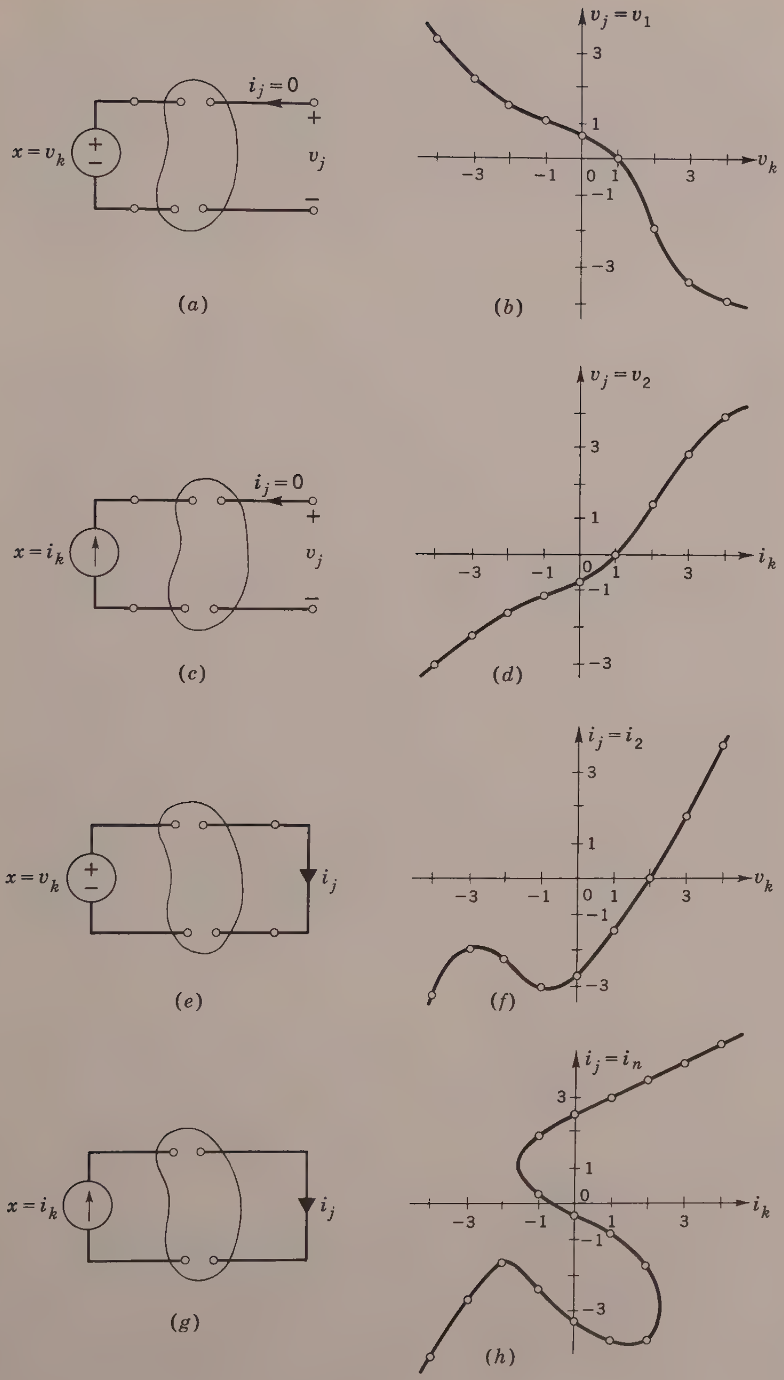


Fig. 5-11. Examples of the four different types of TC plots.

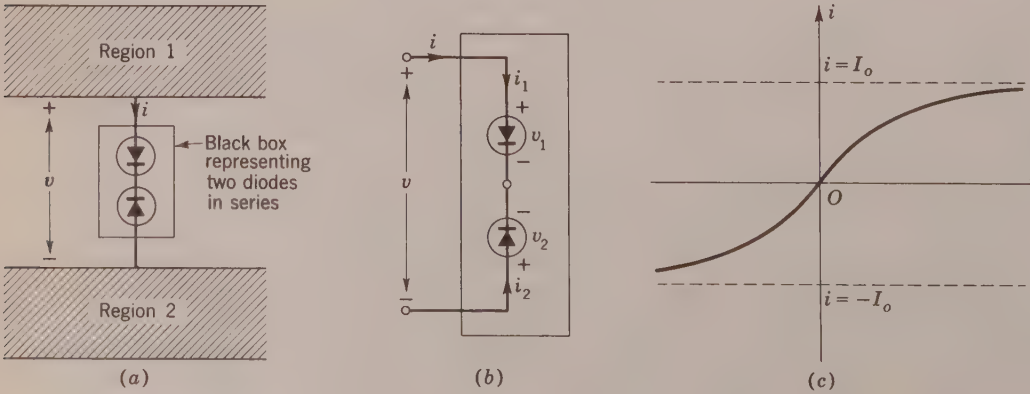
5-3 SOME PRACTICAL EXAMPLES

The concepts of the DP and TC plots rank foremost in importance among all concepts to be introduced in this book. As we shall see, these concepts pop up over and over again in both resistive and dynamic nonlinear networks. Before turning to a rigorous presentation of the analysis (Chaps. 6 and 7) and the synthesis (Chap. 8) of DP and TC plots, let us consider a few simple examples of some of the many applications of these basic concepts.

5-3-1 ISOLATION BLOCK FOR INTEGRATED CIRCUITS

Because the development of integrated-circuit technology was based originally on the mass fabrication of devices such as transistors and diodes on a single "chip" of material, such as silicon, many new practical constraints were created. Among them is the fact that it is more economical to fabricate transistors and diodes than linear resistors. Moreover, for linear resistors the resistance value compatible with present technology is relatively low. Another problem is isolation. Because all devices are, in a way, grown on the same base, known as a *substrate,* it is necessary to provide enough isolation so that each device will behave independently. One practical method to achieve this isolation is to fabricate two junction diodes back-to-back and sandwich them between the two regions to be isolated, as depicted *symbolically* by the drawings shown in Fig. 5-12*a*. If the isolation is perfect, there should be no current flowing from region 1 to region 2, and vice versa. However, a current i does flow when a voltage v exists across the two regions. In order to be able to include this effect in the final design, it is necessary to find the relationship between v and i. If we

Fig. 5-12. The relationship between v and i of the isolation black box is simply the DP plot of the two diodes in series.



enclose the two diodes by a box as shown in Fig. 5-12*b*, then the desired *v*-*i* relationship is, by definition, the DP plot of this black box.

For this simple circuit, it is possible for us to derive the DP plot analytically. Let us represent the diodes by the approximate relationship given earlier by Eq. (1-14); thus

$$\begin{aligned} i_1 &= I_o(e^{kv_1} - 1) \\ i_2 &= I_o(e^{kv_2} - 1) \end{aligned} \tag{5-9}$$

Equations (5-9) represent the laws of the elements. The laws of interconnection are

KVL: $v = v_1 - v_2$ (5-10)

KCL: $i = i_1 = -i_2$ (5-11)

Since our objective is to solve for v as a function of i, or vice versa, an inspection of Eqs. (5-10) and (5-11) suggests that we solve for v_1 and v_2 from Eq. (5-9) in terms of i_1 and i_2; thus

$$\begin{aligned} v_1 &= \frac{1}{k} \ln \left(\frac{i_1 + I_o}{I_o} \right) \\ v_2 &= \frac{1}{k} \ln \left(\frac{i_2 + I_o}{I_o} \right) \end{aligned} \tag{5-12}$$

Substituting Eqs. (5-11) and (5-12) into Eq. (5-10), we obtain, after some simplification,

$$v = \frac{1}{k} \ln \left(\frac{I_o + i}{I_o - i} \right) \tag{5-13}$$

Since we normally plot the current as the vertical axis (by tradition), let us solve for i in terms of v from Eq. (5-13) to obtain

$$i = I_o \frac{e^{kv} - 1}{e^{kv} + 1} \tag{5-14}$$

This is the equation of the DP plot across the two diodes. It can be easily plotted by observing that $i \to I_o$ as $v \to \infty$, and $i \to -I_o$ as $v \to -\infty$. Moreover, this curve is symmetrical with respect to the origin, as shown in Fig. 5-12*c*. In practice, the current I_o is called the *saturation* current and is usually very small, a few microamperes. From this DP plot, we observe that the magnitude of the leakage current i increases with the voltage v and eventually approaches the saturation current I_o. Since I_o is a very small current, it is clear that the isolation is quite good. If we define the dc resistance at each point on the DP plot to be $R_{dc} = v/i$ and the ac resistance to be $R_{ac} = dv/di$, we obtain[1]

$$R_{dc} = \frac{v}{I_o} \frac{e^{kv} + 1}{e^{kv} - 1} \tag{5-15}$$

$$R_{ac} = \frac{2}{k} \frac{I_o}{I_o^2 - i^2} \tag{5-16}$$

It is easy to see that these resistances represent fairly high values, and hence this is one practical method for obtaining high resistances in integrated circuits.

Exercise 1: (*a*) Derive Eqs. (5-12) to (5-16). (*b*) Sketch R_{dc} versus v and R_{ac} versus i.

Exercise 2: Consider the dual topology where the junction diodes are connected in parallel with opposite polarities. (*a*) Show that the DP plot is given by $i = 2I_o \sinh kv$. (*b*) Why is the equation not similar in form to Eq. (5-13)? (*c*) Derive the corresponding expressions for R_{dc} and R_{ac}.

5-3-2 A SQUARE-LAW TRANSFER-CHARACTERISTIC PLOT

The simple circuit shown in Fig. 5-13*a* has a number of very practical applications. In practice, such a circuit is usually designed to perform a certain operation on the input signal $v_i(t)$ to give a desired output signal $v_o(t)$. If we maintain $i_o = 0$, then this operation is completely specified by the v_o-vs.-v_i TC plot because once this relationship (either analytically or graphically) is known, we can obtain the output voltage $v_o(t)$ corresponding to any input voltage $v_i(t)$. For example, if $v_o = A \ln v_i$ and $v_i(t) = 2 + \sin \omega t$, then $v_o(t) = A \ln (2 + \sin \omega t)$. A circuit with the above TC plot is called a *logarithmic amplifier* and is useful in such diverse applications as radar, nuclear instrumentation, and dynamic range extenders.

To derive the TC plot, let us assume that the nonlinear resistor is characterized by $v_1 = f_1(i_1)$ and obtain the equations

$$v_i = f_1(i) + Ri \tag{5-17}$$

and

$$v_o = Ri \tag{5-18}$$

---

[1] See Prob. 1-4 of Chap. 1.

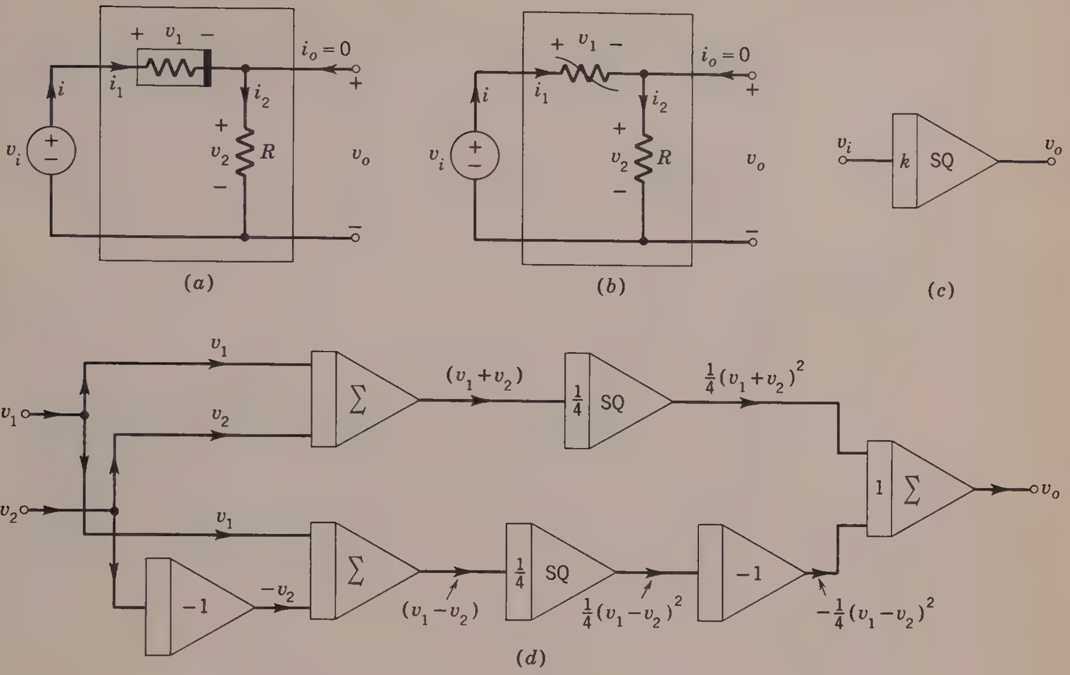


Fig. 5-13. Any circuit with a square-law TC plot can be used as a basic building block for designing a multiplication circuit.

We can eliminate i by substituting Eq. (5-18) into Eq. (5-17) to obtain

$$v_i = f_1\left(\frac{v_o}{R}\right) + v_o \equiv f(v_o) \tag{5-19}$$

This equation specifies the v_o-vs.-v_i TC plot in the v_i-v_o plane. Since we would normally plot the independent variable v_i as the horizontal axis, it is convenient to solve for v_o as a function of v_i from Eq. (5-19) to obtain

$$v_o = g(v_i) \tag{5-20}$$

where $g(v_i) \equiv f^{-1}(v_i)$ is simply the inverse function of $f(v_o)$.† In general, Eq. (5-20) is either multivalued or cannot be so expressed in terms of elementary functions. But in any case, its TC plot can be plotted directly from Eq. (5-19).

As a very common application, let us choose the nonlinear resistor to be a varistor (Fig. 5-13*b*) with its v_1-i_1 curve represented approximately by[1]

$$v_1 = \alpha i_1{}^{\beta} = f_1(i_1) \tag{5-21}$$

---

† Recall that the inverse f^{-1} of a function f exists if, and only if, f is a strictly monotonic function.

[1] In practice, the varistor is usually called a Thyrite, a registered trademark of the Carboloy Department of the General Electric Company.

The practical values of α and β usually lie in the range $10 < \alpha < 5{,}000$, and $0.10 < \beta < 0.45$. Substituting Eq. (5-21) for f_1 in Eq. (5-19), we obtain

$$v_i = \alpha \left(\frac{v_o}{R}\right)^{\beta} + v_o \tag{5-22}$$

The slope of this TC plot is readily obtained to be

$$\frac{dv_i}{dv_o} = \frac{\alpha\beta}{R}\left(\frac{R}{v_o}\right)^{1-\beta} + 1 > 0 \tag{5-23}$$

since $(1 - \beta) > 0$ for $0.10 < \beta < 0.45$. Therefore, this TC plot is a strictly monotonically increasing curve and possesses an inverse in the form of Eq. (5-20). For a given varistor, the exact values of α and β can be measured, and an appropriate value of R can be found by the graphical methods in Chap. 6 so as to best approximate a square-law TC plot[1]

$$\begin{aligned} v_o &= kv_i^2 && \text{for all } v_i \geq 0 \\ &= -kv_i^2 && \text{for all } v_i < 0 \end{aligned} \tag{5-24}$$

This particular TC plot is useful in many applications. Accordingly, we shall denote the corresponding circuit by the symbol shown in Fig. 5-13*c* where both v_i and v_o are measured with respect to ground (not shown for simplicity).[2]

One of the most common applications of a square-law circuit is the multiplication of two voltages $v_1(t)$ and $v_2(t)$. It is based on the identity

$$v_1v_2 = \tfrac{1}{4}(v_1 + v_2)^2 - \tfrac{1}{4}(v_1 - v_2)^2 \tag{5-25}$$

With the help of the summing and phase-inverting components presented in Sec. 3-6-2, it is easy to see that the circuit shown in Fig. 5-13*d* constitutes a realization of the relationship

$$v_o = v_1v_2 \tag{5-26}$$

Hence, in addition to the list of basic mathematical operations discussed in Sec. 3-6-2, we can now also perform multiplication.

Exercise 1: (*a*) Show that the logarithmic amplifier can be realized by the circuit shown in Fig. 5-13*a* if the nonlinear resistor is replaced by a resistive black box with the DP plot $v_1 = e^{ki_1} - i_1R$, where $k = R/A$. (*b*) Considering the above procedure, discuss the general method for finding an appropriate DP plot in order to realize a prescribed TC plot.

---

[1] For more practical details on the design of this TC plot, see L. D. Kovach and W. Comley, Nonlinear Transfer Functions with Thyrite, *IRE Trans. Electron Computers*, pp. 91–97, June, 1958.

[2] A true square-law circuit is characterized by $v_o = kv_i^2$ for all v_i. Although there are other circuits with this TC plot, for simplicity we shall use our present circuit with $v_i \geq 0$.

Exercise 2: Derive a general v_o-vs.-v_i TC plot relationship analogous to Eq. (5-19) if the two resistors in Fig. 5-13*a* are interchanged.

Exercise 3: Explain why it is that the two square-law circuits may be connected as shown in Fig. 5-13*d* without introducing loading effects.

5-4 THE POWER-TRANSFER PLOT

Any physical device must be operated below its maximum power rating. For a dc-resistive network, the power dissipated in each device can be found by determining first the operating points of the network. In the case of a two-terminal nonlinear resistor, the operating point must be restricted between the maximum-power hyperbola $vi = P_{\max}$, where $P_{\max}$ is the maximum power rating of the device, and the axes. For example, the operating point on the varistor v-i curve shown in Fig. 5-14*a* must be restricted from the shaded region. In the case of a three-terminal nonlinear resistor,

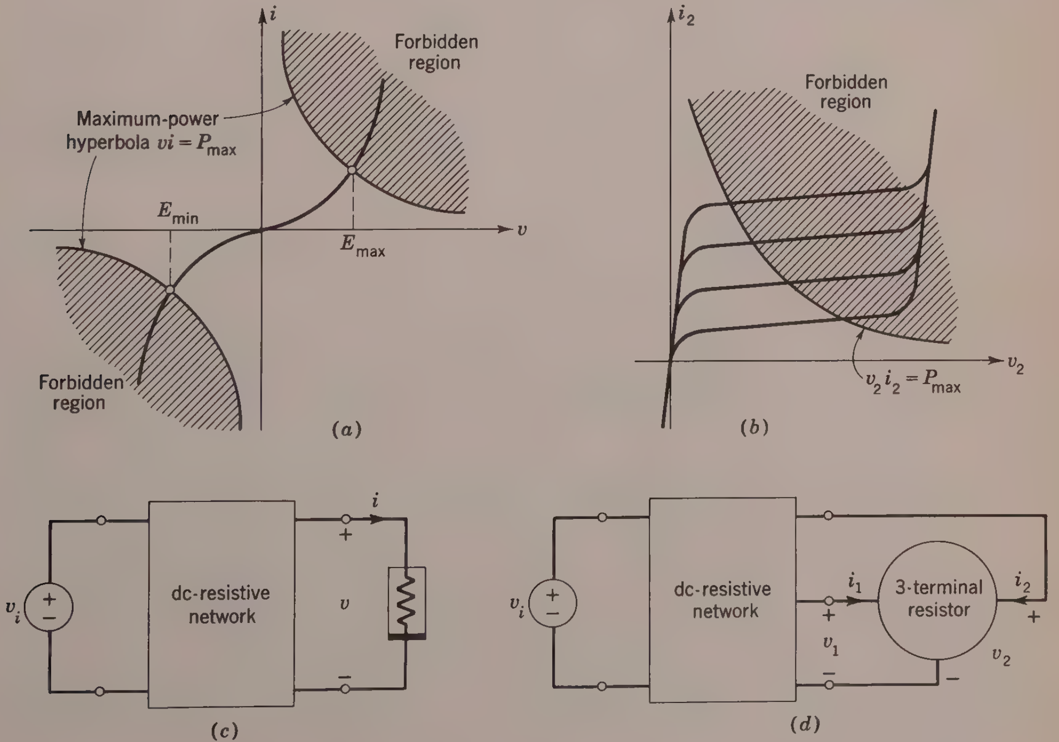


Fig. 5-14. The power dissipated in any two-terminal or three-terminal device must lie under the maximum-power hyperbola at all times.

the total input and output power $P_T = v_1 i_1 + v_2 i_2$ must not exceed the maximum rating. In many practical devices, such as the FET, the input power $v_1 i_1$ is usually negligible and hence $P_T \approx v_2 i_2$. For example, the operating point on the drain-to-source characteristic curves of the FET shown in Fig. 5-14*b* must lie between the maximum-power hyperbola $v_2 i_2 = P_{\max}$ and the axes i_2 and v_2.

For ac networks, the operating point varies with time, and hence it becomes rather impractical to solve for the operating points for all times. Fortunately, most practical ac networks of interest consist of a dc network driven by a single ac independent voltage source. In this case, it is more convenient for us to solve for the maximum range of voltages that can be applied without exceeding the ratings of the elements in the network. In most cases, the elements of interest are usually two-terminal or three-terminal nonlinear resistors. For the sake of discussion, let us extract the nonlinear resistor under consideration as shown in Fig. 5-14*c* and *d*.

Consider first the case of the two-terminal resistor. The power dissipated in this resistor is given by $p = vi$. If we solve for the v-vs.-v_i and the i-vs.-v_i TC plots

$$v = f(v_i)$$
$$i = g(v_i)$$

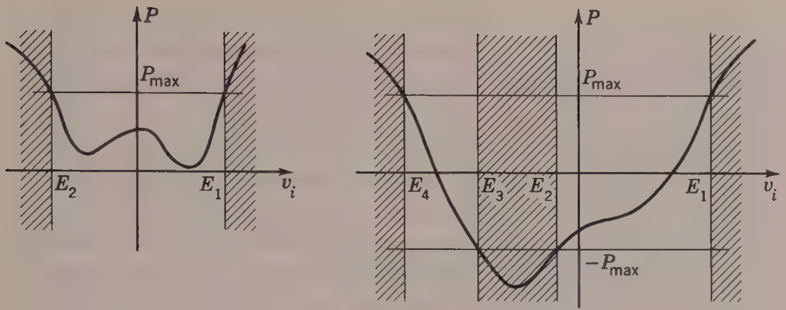
we obtain

$$p = f(v_i)g(v_i) \equiv h(v_i) \tag{5-27}$$

This relationship represents the power dissipated in the nonlinear resistor in Fig. 5-14*c* as a function of the ac voltage v_i, and is called the *power-transfer plot*. Once this relationship is found, the range of the input voltage that can be applied is easily obtained. If the resistor is *passive*, then $P \geq 0$ for all v_i as shown by the typical power-transfer plot in Fig. 5-15*a*. In this case, the range of allowable voltage is given by $E_2 < v_i < E_1$. If the resistor is *active*, then P may be negative as shown by the typical power-transfer curve in Fig. 5-15*b*. In this case, there are two nonoverlapping ranges of permissible voltage, namely, $E_2 < v_i < E_1$ and $E_4 < v_i < E_3$.

Consider next the three-terminal resistor case shown in Fig. 5-14*d*. The power-transfer plot is now defined by

$$P = f_1(v_i)g_1(v_i) + f_2(v_i)g_2(v_i) \tag{5-28}$$



(a) (b)

Fig. 5-15. Typical examples of the power-transfer plot of (a) passive and (b) active nonlinear resistors.

where $v_1 = f_1(v_i)$, $i_1 = g_1(v_i)$, $v_2 = f_2(v_i)$, and $i_2 = g_2(v_i)$ are the TC plots from the respective terminal variables of the three-terminal resistor to the driving-point voltage. In many practical three-terminal resistors, the power entering port 1 of the resistor is negligible, in which case Eq. (5-28) reduces to

$$P \approx f_2(v_i)g_2(v_i) \tag{5-29}$$

The utility of the power-transfer plot is now obvious; it gives the designer a complete picture of how much power is dissipated in a nonlinear resistor corresponding to any value of the applied input voltage. Equations (5-27) and (5-28) show that this power-transfer plot can be found by multiplying appropriate TC plots.[1] This is another occasion where the concept of TC plots is useful.

Exercise 1: Derive similar relationships for the power-transfer plot for the case where the ac-voltage source is replaced by an ac-current source.

Exercise 2: In most practical circuits, the TC plots must be found in graphical form. Give a simple, graphical procedure for obtaining the power-transfer plot.

Exercise 3: Show that if the v-i curve of the nonlinear resistor in Fig. 5-14c is characterized by a current-controlled curve $v = v(i)$, then the power-transfer plot defined by Eq. (5-27) can be found from this v-i curve and the $i = g(v_i)$ TC plot. Give a graphical construction procedure. HINT: Use the graphical composition technique given in Appendix B.

Exercise 4: Show that the power-transfer plot in Exercise 3 can be found directly from the v-i curve $i = i(v)$ of a voltage-controlled resistor and the $v = f(v_i)$ TC plot. Give a graphical construction procedure.

5-5 THE OPERATING-POINT PARADOX

It was mentioned in Sec. 5-1 that a resistive network may have one operating point, a finite number of operating points, an infinite number of operating points, or no operating point at all.

---

[1] Actually, only one TC plot needs to be found. See Exercises 3 and 4.

This statement seems to contradict what we observe physically, namely, that any physical network can have one, and only one, operating point at any one time. Something must be wrong somewhere!

Actually, there is nothing wrong in both statements above. The paradox can be resolved upon recognizing that no physical network is purely resistive. Some parasitic inductances in the wires and some parasitic capacitances between conductors are invariably associated with any physical network. Consequently, when we consider a resistive network, we are actually dealing with a nonphysical network. We have no reason, therefore, to expect a nonphysical network to yield physical answers.

But why study a nonphysical network? The reason is simple: as will be shown in Part 3 of this book, the analysis of any dynamic network *N* invariably requires a preliminary analysis of a resistive subnetwork *N′* obtained from the former by replacing all inductors in *N* by short circuits and all capacitors in *N* by open circuits. The fact that *N′* may have more than one operating point corresponds to the physically observed fact that a network may be forced to operate at one of the several possible operating points at any one time. For example, a well-known circuit called the flip-flop is a dynamic network which can be made to operate at one of two possible states. It will be shown in Chap. 13 that these two states can be found by solving for the operating points of the associated resistive subnetwork. This observation alone justifies a detailed study of resistive networks. As a matter of fact, the reader will soon discover for himself that a mastery of the basic concepts in nonlinear resistive networks is crucial to a complete understanding of dynamic networks. The study of nonlinear resistive networks therefore serves as a means to an end, rather than the end itself.

5-6 THE THREE FUNDAMENTAL THEOREMS OF RESISTIVE NETWORKS

We have presented here three fundamental concepts, the operating-point concept, the DP plot concept, and the TC plot concept. We have demonstrated that at any time *t*, these concepts are well-defined and meaningful for any solvable resistive network.[1] The network may contain any number of *n*-terminal resistors, independent sources, ideal transformers, gyrators, and controlled sources. In other words, no matter how complicated the network is, we can always talk about its operating point at any given time *t*. Moreover, if we pick any one of the dc sources in a dc-resistive

[1] A network is solvable if it has at least one solution. We shall exclude unsolvable networks in this book because they are only of academic interest.

network and vary its terminal voltage or current, we can plot the resulting current (through an arbitrarily chosen wire) or the voltage (across an arbitrarily chosen pair of terminals) as a function of the terminal current or voltage of the dc source. This means that no matter how complicated the dc network is, we can always talk about the DP plot across a pair of driving-point terminals created either by a soldering-iron entry or by a pliers-type entry. We can also talk about the TC plot between a driving-point variable and an output variable, where either variable may be a current or a voltage. For ac networks, the concepts of DP and TC plots are still valid, but they will vary with time. The crucial property that ensures the validity of these concepts is the fact that the equations of motion of any resistive nonlinear network is always a system of nonlinear functional equations. In view of their fundamental importance, we shall summarize the above results as two separate theorems.

THE FIRST FUNDAMENTAL THEOREM OF RESISTIVE NETWORKS, THE DP PLOT THEOREM

At any time t, a DP plot always exists across any pair of driving-point terminals created (either by a soldering-iron entry or by a pliers-type entry) across any solvable resistive nonlinear network. This DP plot can be determined either by the methods to be presented in the following chapters or by actual measurements. Moreover, in the case of a dc network, the DP plot can be interpreted as the v-i curve of an equivalent two-terminal resistor obtained by considering the entire network as a two-terminal black box.

THE SECOND FUNDAMENTAL THEOREM OF RESISTIVE NETWORKS, THE TC PLOT THEOREM

Given any solvable resistive nonlinear network and an input variable representing either a voltage source or a current source across some driving-point terminals, then at any time t a TC plot always exists relating an arbitrary output variable, which may be either the current through some wire or the voltage across some pair of terminals, as a function of the input variable. This TC plot can be determined either by the methods to be presented in the following chapters or by actual measurements.

Another fundamental result that we shall find much use for later is the uniqueness theorem. Since a complete proof of this theorem is too complicated to be given in an introductory book of this nature, we shall only state it here.

THE THIRD FUNDAMENTAL THEOREM OF RESISTIVE NETWORKS, THE UNIQUENESS THEOREM

Let N be a resistive network containing only two-terminal resistors and independent sources (no controlled sources). If the v-i curve of each resistor is a strictly monotonically increasing function which tends to $\pm\infty$ as v tends to $\pm\infty$, and if N contains neither loops of voltage sources nor cut sets of current sources, then the following statements are always true:

1. Regardless of the instantaneous values of all independent (ac and dc) voltage and current sources, the network N has one, and only one, solution for all time t.
2. The DP plot across any pair of driving-point terminals is always a strictly monotonically increasing curve.
3. The magnitude of the slope at any point on a v_2-vs.-v_1 TC plot or an i_2-vs.-i_1 TC plot is always less than or equal to 1.†

5-7 SUMMARY

Three basic concepts of resistive nonlinear networks

1. The operating point is simply the solution of a resistive nonlinear network.
2. The driving-point characteristic plot (DP plot) gives the relationship between the driving-point current i and driving-point voltage v. It can be interpreted as the v-i curve of a black box.
3. The transfer-characteristic plot (TC plot) gives the relationship between the voltage across (or the current through) a pair of transfer terminals and the voltage v or the current i across a pair of driving-point terminals. Accordingly, there are four types of TC plots:
   a. v_2-vs.-v_1 plot
   b. i_2-vs.-i_1 plot
   c. v_2-vs.-i_1 plot
   d. i_2-vs.-v_1 plot

Power-transfer plot This plot gives the relationship between the power P dissipated in a nonlinear resistor and the driving-point voltage v or the driving-point current i. The power-transfer plot can be determined by multiplying an appropriate TC plot and the resistor v-i curve.

---

† This property is generally not true for the remaining two types of TC plots.

First fundamental theorem A DP plot exists across any driving-point terminals of a resistive nonlinear network.

Second fundamental theorem A TC plot exists between any pair of transfer terminals and driving-point terminals.

Third fundamental theorem Any resistive network containing only strictly monotonically increasing resistors (such that $|i| \to \infty$ as $|v| \to \infty$) possesses (1) a unique solution, (2) a strictly monotonically increasing DP plot across any driving-point terminals, and (3) a v_2-vs.-v_1 or i_2-vs.-i_1 TC plot with slopes less than or equal to unity.

PROBLEMS

5-1 Consider the network shown in Fig. P5-1 where the nonlinear resistor is characterized by

$$i_4 = 0 \qquad v_4 \leq 0$$

$$i_4 = v_4^2 \qquad v_4 > 0$$

(a) Find the operating point; that is, find v_j and i_j for all $j = 1, 2, \ldots, 5$.

(b) If the sources in Fig. P5-1 are replaced by

$$v_1 = 17 - 5 \cos t$$

$$v_2 = 1⅔ \cos t + ⅓ \cos 2t$$

find the solution $v_j(t)$ and $i_j(t)$ for all $t \geq 0$ and for all $j = 1, 2, \ldots, 5$.

(c) Compare your answers for (*a*) and your answers for (*b*) at $t = 0$. Based on this observation, explain how an ac-resistive nonlinear network can be analyzed as a number of operating-point problems at various instants of time.

(d) On what occasions will your conclusion in (*c*) be useful?

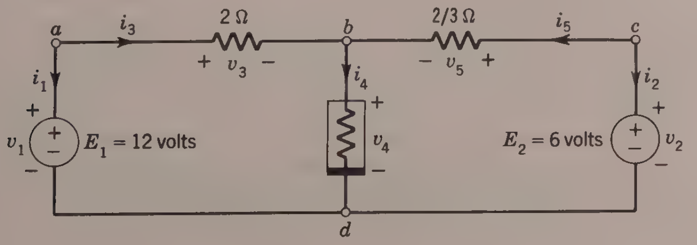


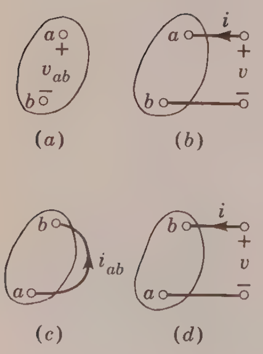
Fig. P5-1.

5-2 For the network shown in Fig. P5-1:
- (a) Find the DP plot across the terminal pair formed by a soldering-iron entry across terminals *a-b*. Assume terminal *a* to be the positive terminal.
- (b) Repeat (*a*) across terminals *b-d* with terminal *b* chosen to be the positive terminal.
- (c) Find the DP plot across the terminal pair formed by a pliers-type entry through the wire connecting terminal *b* with the ⅔-Ω resistor. Assume terminal *b* to be the positive terminal.
- (d) Repeat (*c*) through the wire connecting terminal *d* with the nonlinear resistor. Assume terminal *d* to be the positive terminal.

5-3 For the network shown in Fig. P5-1:
- (a) Find the TC plot between the voltage v_3 and the voltage v across the pair of driving-point terminals formed by pliers-type entry between terminal *b* and the nonlinear resistor. Assume terminal *b* to be the positive terminal.
- (b) Find the TC plot between the current i_4 and the voltage v across the pair of driving-point terminals formed by pliers-type entry between terminal *c* and the ⅔-Ω resistor. Assume terminal *c* to be the positive terminal.
- (c) Find the TC plot between the current i_4 and the current i entering terminal *b* of the driving-point terminals formed by a soldering-iron entry across terminals *b-c*.
- (d) Find the TC plot between the voltage v_3 and the current i entering terminal *b* of the driving-point terminals formed by a soldering-iron entry across terminals *b-d*.

5-4 (a) Let v_{ab} be the voltage between any pair of nodes *a-b* of a dc-resistive nonlinear network N as shown in Fig. P5-4*a*. Let v be the voltage across the driving-point terminals formed by a soldering-iron entry across nodes *a-b* as shown in Fig. P5-4*b*. Prove that the voltage v_{ab} can be found by determining the voltage at the point of intersection between the DP plot i versus v with the v axis.
- (b) Find the values of v_3 and v_4 associated with the network in Fig. P5-1 by applying the method given in (*a*) to the DP plots obtained in Prob. 5-2*a* and *b*.
- (c) Let i_{ab} be the current in any wire of a dc-resistive nonlinear network N as shown in Fig. P5-4*c*. Let i be the current entering the driving-point terminals formed by a pliers-type entry through the wire as shown in Fig. P5-4*d*. Prove that the current i_{ab} can be found by determining the current at the point of intersection between the DP plot i versus v with the i axis.
- (d) Find the values of i_5 and i_4 associated with the network in Fig. P5-1 by applying the method given in (*c*) to the DP plots obtained in Prob. 5-2*c* and *d*.

Fig. P5-4.



5-5 (a) For the network shown in Fig. P5-5 determine and sketch the TC plot v_o versus v_1.

(b) If $v_1(t) = 8 \sin \omega t$, find $v_o(t)$ using the TC plot obtained in (*a*). Interpret your procedure graphically.

(c) If an R-ohm linear resistor is connected across terminals *c-d* of the network in Fig. P5-5, find the TC plot v_o versus v_1 in terms of R. Sketch the TC plots for $R = 0, 2, 10, \infty$. Interpret the loading effect due to R. Can you use the TC plot found in (*a*) as an aid in finding these TC plots?

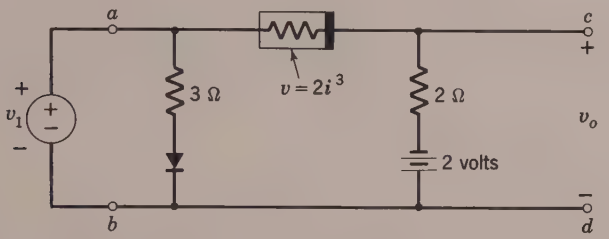


Fig. P5-5.

5-6 Consider the resistive nonlinear network shown in Fig. P5-6*a*. Resistor R_1 is described by $v_1 = -\frac{1}{2}$ for $i_1 \leq -1$ and $v_1 = \frac{1}{2}i_1^2$ for $i_1 > -1$. Resistor R_2 is described by $v_2 = 3i_2^2$.

(a) Derive and sketch the DP plot i versus v if I_s is a current-controlled current source given by $I_s = 5i$.

(b) Repeat (*a*) for values of I_s equal to 2 and 4.

(c) Repeat (*a*) if I_s is an ac-independent current source described by the time function shown in Fig. P5-6*b*.

(d) Repeat (*c*) if $I_s(t) = 2t$.

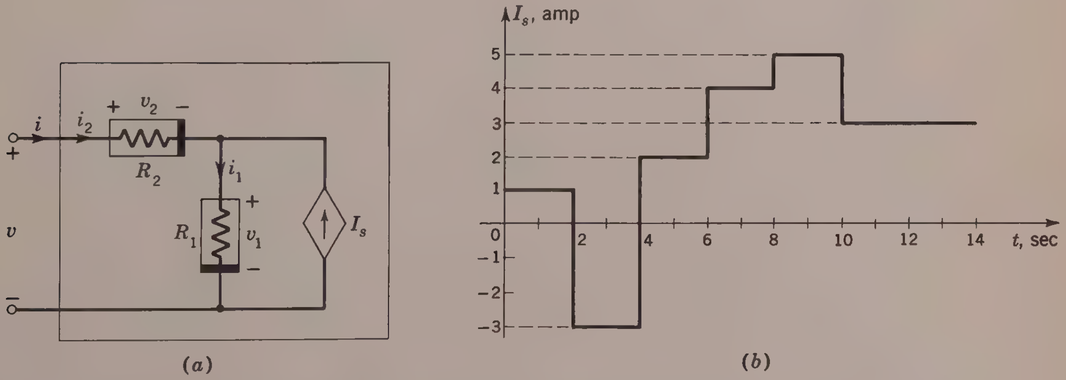


Fig. P5-6.

5-7 Consider the network shown in Fig. P5-7*a* where the nonlinear resistor is characterized by $i = \frac{1}{17}v^3$.

(a) Find the TC plot v_5 versus v_1.

(b) Find the TC plots v_6 versus v_1 and v_2 versus v_1.

(c) Consider the network configuration shown in Fig. P5-7*b*. Suppose the TC plots v_{ca} versus v_1 and v_{bd} versus v_1 are given. Show how to determine the TC plot v_{cd} versus v_1.

(d) Repeat (*a*) using the method given in (*c*) and the TC plots obtained in (*b*).

HINT: Consult any standard mathematical handbook for a formula giving the solutions of a cubic equation.

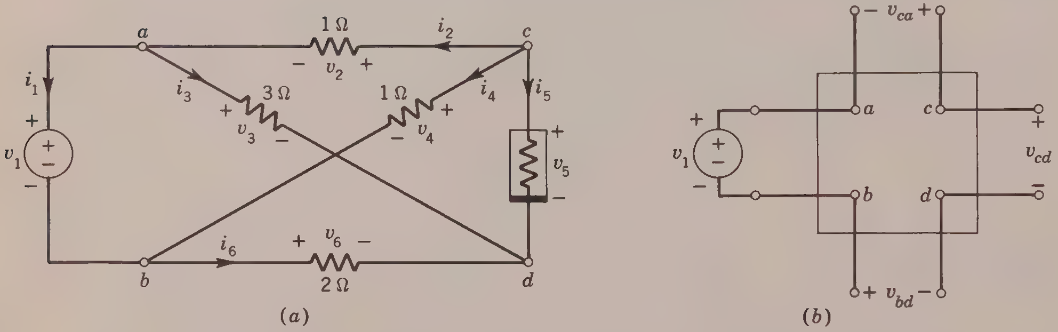


Fig. P5-7.

5-8 For the network in Fig. P5-8, the three nonlinear resistors are characterized by

$$i_1 = f_1(v_1) = v_1^2 + 2v_1$$

$$i_2 = f_2(v_2) = 5v_2 - v_2^3$$

$$i_3 = f_3(v_3) = v_3^2$$

The three-terminal resistor is characterized by

$$i_4 = f_4(v_4,v_5) = v_5^2 - 2v_4v_5 + 2v_4^2 + 5$$

$$i_5 = f_5(v_4,v_5) = v_5^3 - v_5^2 - v_4^2 + 10$$

(a) Find the operating point with $v = 0$.
(b) Find the DP plot i versus v across terminals *a-b*.
(c) Find the TC plot of v_2 versus v.

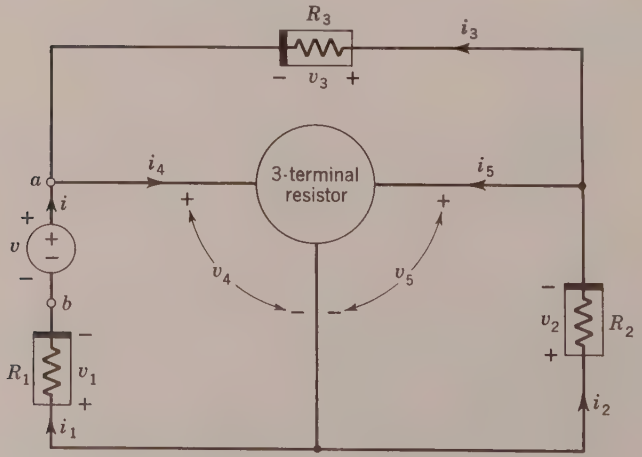


Fig. P5-8.

5-9 For the network in Fig. P5-9, the two nonlinear resistors are characterized by

$$v_1 = f_1(i_1) = \tfrac{3}{4}i_1{}^3 - \tfrac{10}{3}i_1$$

$$i_2 = f_2(v_2) = \tfrac{1}{2}v_2{}^3 - \tfrac{11}{2}$$

The three-terminal resistor is characterized by

$$v_4 = f_4(i_4,v_5) = -5i_4 + \tfrac{5}{2}v_5 + \tfrac{1}{3}v_5{}^3$$

$$i_5 = f_5(i_4,v_5) = -10i_4 + \tfrac{1}{2}v_5{}^3 + 2$$

(a) Find the operating point with $v = 0$.

(b) Find the DP plot across terminals a-b analytically; i.e., find a relationship of the form $f(i,v) = 0$.

(c) Find the TC plot of v_2 versus v.

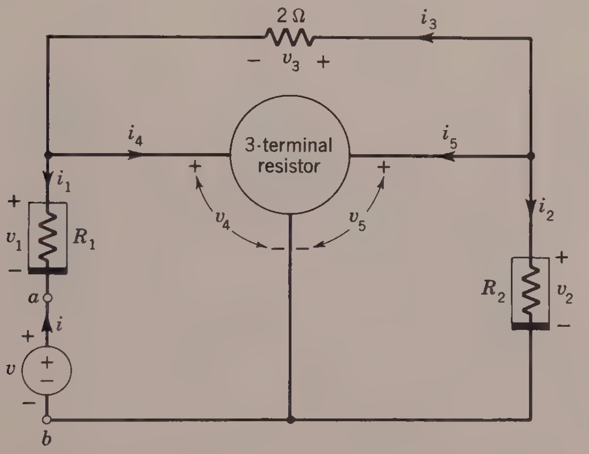


Fig. P5-9.

5-10 Consider the network in Fig. P5-10 where the nonlinear resistor R_3 is characterized by $v_3 = 2i_3{}^3$.

(a) Find the TC plot i_3 versus i.

(b) Let P_3 be the power dissipated in resistor R_3. Use the relationship found in part (a) and the characteristic of R_3 to find the power-transfer plot P_3 versus i.

HINT: Consult a mathematical handbook for a formula giving the solutions to a cubic equation.

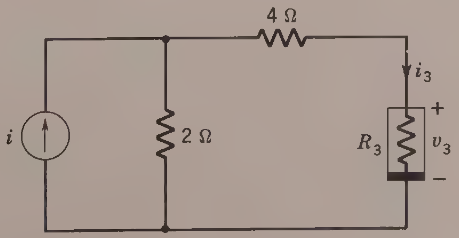


Fig. P5-10.

5-11 Consider the network shown in Fig. P5-11, where the nonlinear resistors are characterized as follows:

R_1: $v_1 = 2i_1^{1/3}$

R_2: $i_2 = 4v_2^3 - 2v_2$

R_3: $i_3 = 4v_3^3 + 2v_3$

The maximum allowable power dissipation for R_3 is 100 watts.

(a) Analytically find the power-transfer plot P_3 versus v. Determine the allowable range for v.

(b) Find the TC plot v_3 versus v. Use this TC plot and the DP plot of R_3 to find P_3 versus v graphically.

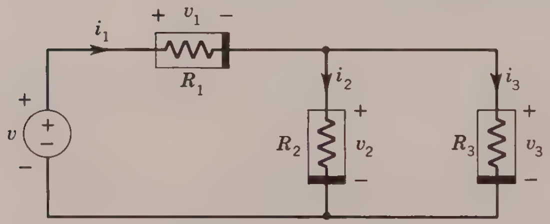


Fig. P5-11.

5-12 The network in Fig. P5-12 contains two independent sources and a nonlinear resistor characterized by $v = 3i^3 - \frac{3}{4}i$. Find the power dissipated in the 1-Ω resistor as a function of v_1 and v_2.

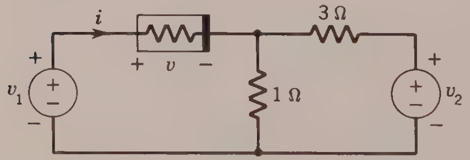


Fig. P5-12.

6 GRAPHICAL ANALYSIS OF RESISTIVE NONLINEAR NETWORKS

6-1 A BIRD'S-EYE VIEW

In this chapter, we shall present a unified *graphical approach* for determining the operating points, the DP plots, and the TC plots of resistive nonlinear networks. We shall first consider series-parallel networks containing only two-terminal resistors and dc independent sources. By a series-parallel network, we mean a network made up of successive series and parallel combinations of two or more two-terminal elements connected in series or in parallel. When a given network contains only a few elements (say less than 10), we can usually determine by inspection whether it is series-parallel. However, when the network is complicated, a systematic test is desirable. One simple test consists of drawing the graph of the network and then replacing each set of parallel branches and each set of series branches of the graph by a single branch. The resulting graph would contain fewer branches, some of which may again be in series or in parallel. These branches are replaced by one branch as before. This procedure is repeated as many times as possible. If the final graph reduces to a single branch, the network is series-parallel.

To illustrate the above test, consider the network shown in Fig. 6-1*a* and its graph in Fig. 6-1*b*. The sets of branches $\{1,2\}$ and $\{5,6\}$ are in parallel and are replaced by branches α and β, respectively, in Fig. 6-1*c*. Similarly, the sets of branches $\{3,4\}$ and $\{7,8\}$ are in series and are replaced by branches γ and δ, respectively, in Fig. 6-1*c*. From this simplified graph, we observe that the set of branches $\{\alpha,\gamma\}$ is in parallel and the set of branches $\{\beta,\delta\}$ is in series. Accordingly, they are replaced by α' and β', respectively, in Fig. 6-1*d*. This leads to the single-branch graph shown in Fig. 6-1*e*, and hence the network is series-parallel. Suppose now we move the resistor R_6 from node c into node b as shown

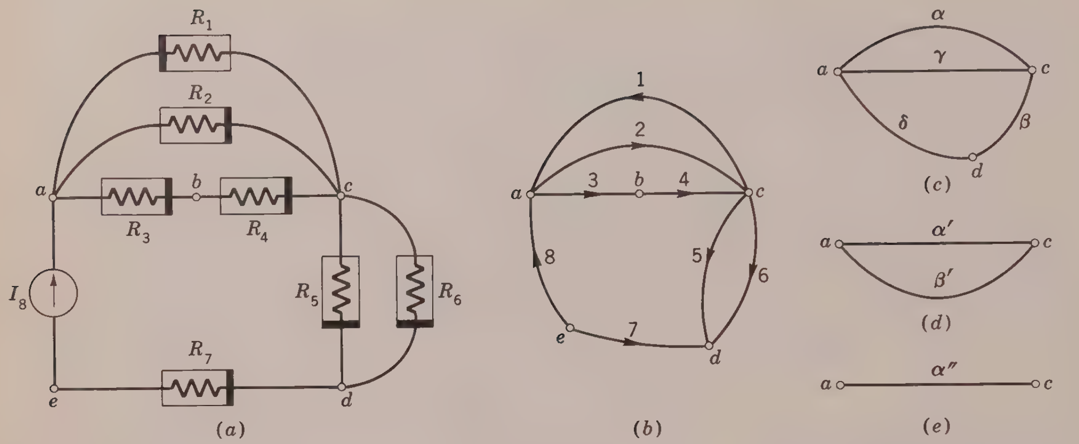


Fig. 6-1. If each set of series branches and each set of parallel branches of a series-parallel network is replaced repeatedly by a single branch, the resulting graph must reduce to one branch.

in Fig. 6-2*a*. The resulting graph shown in Fig. 6-2*b* can be reduced to that shown in Fig. 6-2*c*, but no further. Hence, this network is non-series-parallel.[1]

The class of series-parallel networks is important not only because many practical networks belong to it, but also because it is especially amenable to a graphical analysis. We shall present several simple, graphical procedures for determining the operating points, the DP plots, and the TC plots of any series-parallel networks. These basic procedures will then be generalized to analyze an important class of non-series-parallel networks. In particular, we shall formulate a number of graphical procedures for analyzing a subclass of nonlinear resistive networks containing two-terminal resistors, three-terminal resistors, and dc independent sources. However, we shall not include controlled sources in this chapter, because so far, no general *graphical* procedure is known for analyzing nonlinear networks containing controlled sources.

Exercise 1: Classify each of the following networks according to whether it is series-parallel or not: (*a*) the network in Fig. 4-2*a*; (*b*) the network in Fig. 4-4*a*; (*c*) the network in Fig. 4-7*a*.

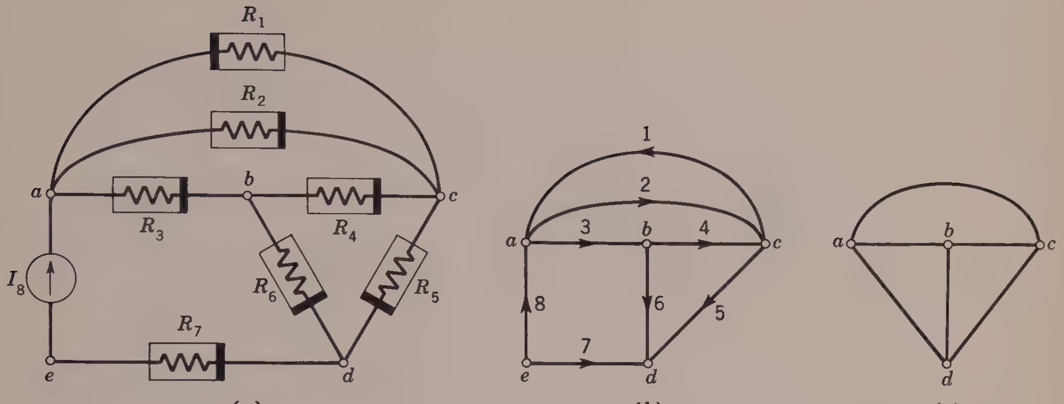
Exercise 2: (*a*) Show that any network of two-terminal elements containing no more than three nodes is series-parallel. (*b*) Show that any "ladder" network is series-parallel.

6-2 GRAPHICAL DETERMINATION OF THE OPERATING POINT

In order to determine the operating point, we must solve for the set of voltages and currents which simultaneously satisfies the equations of motion of a given dc-resistive nonlinear network. This task in concept consists of systematically eliminating one or more variables from the system of equations of motion, and

---

[1] A network which contains one or more three-terminal elements is also non-series-parallel because the terms "series" and "parallel" are meaningful only for interconnections between two-terminal elements.



(a) (b) (c)

Fig. 6-2. An example showing how a non-series-parallel network fails the test.

eventually reducing the system of equations into one equation in one unknown. However, as was pointed out in Chap. 5, the equations of motion of most practical dc-resistive nonlinear networks usually contain functions specified graphically in the form of curves. Hence, the approach described above is not directly applicable. Instead, we apply the graphical approach, which is actually the implementation of the conceptual approach in graphical form. In this section, we shall present several graphical procedures for determining the operating points of two basic network configurations. Most of these procedures are used to carry out graphically the elimination and substitution operations required in the solution of nonlinear equations.

6-2-1 BASIC NETWORK CONFIGURATION 1, INTERCONNECTION BETWEEN 2 TWO-TERMINAL RESISTORS

The simplest network configuration involving two nonlinear resistors is shown in Fig. 6-3*a*. One of the two resistors is commonly referred to as the load. For convenience, we shall usually refer to the resistor with the simpler[1] *v-i* curve as the load. For example, suppose resistors R_1 and R_2 are characterized, respectively, by the curves shown in Fig. 6-3*b* and *c*. Then R_2 will be referred to as the load since its *v-i* curve is simpler than that for resistor R_1. To find the operating point of this network, we must first write the following equations of motion:

Equations from the laws of elements[2]

$$f_1(i_1,v_1) = 0 \tag{6-1a}$$

$$f_2(i_2,v_2) = 0 \tag{6-1b}$$

---

[1] A curve is said to be simpler than another if it can be approximated by fewer piecewise-linear segments.

[2] In order to include both current- and voltage-controlled resistors, as well as multivalued resistors, we have chosen the notation $f_1(i_1,v_1)=0$ and $f_2(i_2,v_2)=0$ to denote the *v-i* curves of R_1 and R_2, respectively.

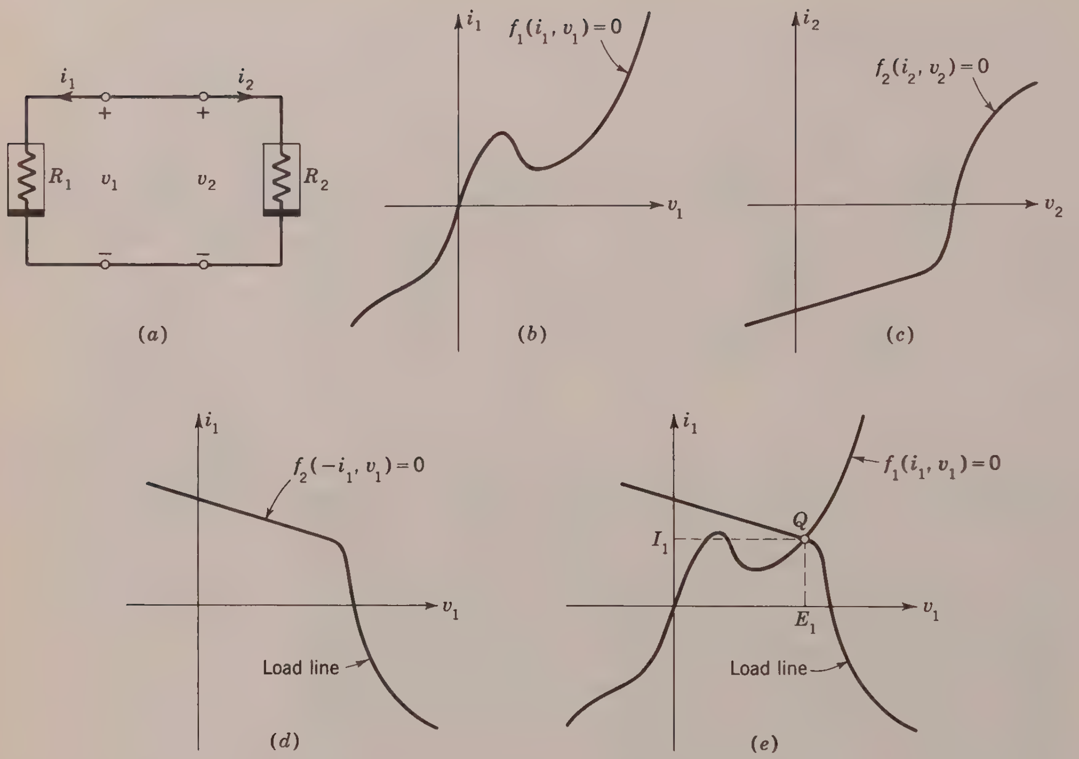


Fig. 6-3. The graphical procedure for determining the operating point of a resistive nonlinear network in basic configuration 1 by the load-line method.

Equations from the laws of interconnection

$$v_2 = v_1 \tag{6-1c}$$

$$i_2 = -i_1 \tag{6-1d}$$

Substituting Eqs. (6-1*c*) and (*d*) into Eqs. (6-1*a*) and (*b*), we obtain

$$f_1(i_1, v_1) = 0 \tag{6-2a}$$

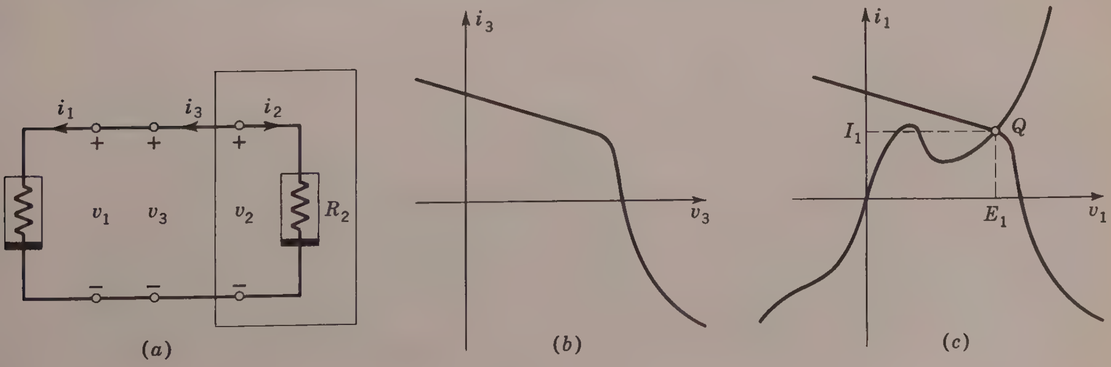
$$f_2(-i_1, v_1) = 0 \tag{6-2b}$$

The operating point can be found by solving for i_1 and v_1 from these equations. Equation (6-2*a*) is still the v_1-i_1 curve shown in Fig. 6-3*b*. However, Eq. (6-2*b*) represents a curve related to the v_2-i_2 curve in Fig. 6-3*c* but with the current coordinates reversed in sign. Hence, Eq. (6-2*b*) represents a curve obtained by reflecting the v_2-i_2 curve in Fig. 6-3*c* with respect to the v_2 axis, as shown in Fig. 6-3*d*. Since the operating point must satisfy both Eqs. (6-2*a*) and (*b*), it can be found by superimposing the curve $f_2(-i_1,v_1) = 0$ (Fig. 6-3*d*) on top of the curve $f_1(i_1,v_1) = 0$ (Fig. 6-3*b*). The point where the curves intersect is the operating point. In Fig. 6-3*e*, this point is labeled Q and has the coordinates (E_1,I_1) in the v_1-i_1 plane. Once E_1 and I_1 are found, the remaining coordinates can be determined from Eqs. (6-1*c*) and (*d*), namely, $i_2 = -I_1$ and $v_2 = E_1$. Thus, the operating point is completely specified by the four values $(E_1,E_1,I_1,-I_1)$. The reflected curve shown in Fig. 6-3*d* is usually called the *load line,* because it is obtained from the v-i curve of the load. Accordingly, the above method is often referred to as the *load-line method.*

There is an alternate and perhaps more intuitive way of interpreting the load-line method. Instead of substituting Eqs. (6-1*c*) and (*d*) into Eq. (6-1*b*), let us enclose R_2 by a black box and define the terminal variables i_3 and v_3 as shown in Fig. 6-4*a*. Since $i_3 = -i_2$ and $v_3 = v_2$, the v_3-i_3 curve of the black box must be as shown in Fig. 6-4*b*. Observe that this curve is identical with the load line in Fig. 6-3*d*. Since the laws of interconnection for this network are given by $i_1 = i_3$ and $v_1 = v_3$, the operating point is simply the intersection between the v_1-i_1 curve of R_1 and the v_3-i_3 curve of the black box, as shown in Fig. 6-4*c*. In both cases, we obtain the same answers, as they should be. These two methods are obviously applicable to any network which can be redrawn as an interconnection between two black boxes in basic configuration 1, provided the DP plot of each black box is given or can be determined.

EXAMPLES

1. Consider the simple switching circuit shown in Fig. 6-5*a* and the tunnel diode v_1-i_1 curve shown in Fig. 6-5*b*. The black box



(*a*) (*b*) (*c*)

Fig. 6-4. An alternate method for finding the operating point of a network in basic configuration 1.

representing the resistor-battery combination is represented by the DP plot

$$v_2 = Ri_2 + E = 500i_2 + 0.45$$

The load line is obtained by substituting $i_2 = -i_1$ and $v_2 = v_1$ in this equation to obtain

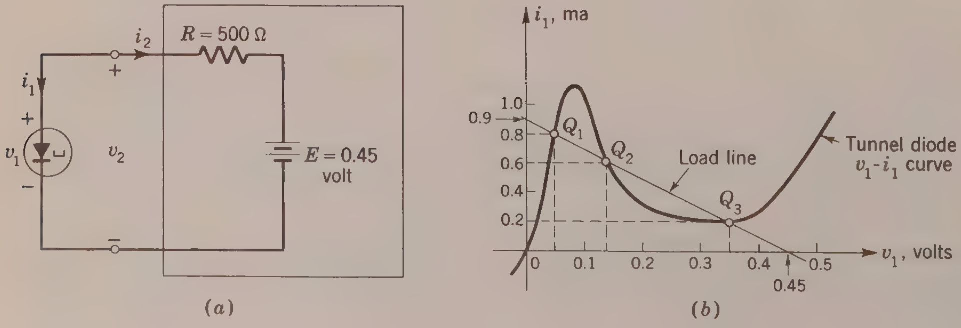
$$v_1 = -(500)i_1 + 0.45$$

The load line corresponding to this equation is drawn on top of the v_1-i_1 curve in Fig. 6-5*b*. Observe that there are three operating points, namely, Q_1:(v_1=0.05 volt, i_1=0.8 ma); Q_2:(v_1= 0.135 volt, i_1 = 0.6 ma); and Q_3:(v_1 = 0.35 volt, i_1 = 0.2 ma).

Since the load consisting of a linear resistor R in series with a battery E is quite common in practice, it is worthwhile to be able to draw the corresponding load line by inspection. This can be done by remembering that the voltage intercept is equal to E, and the current intercept is equal to E/R. Observe that the effect of changing the value of the voltage E is simply to move the load line parallel to itself, and the effect of changing the value of R is to change the slope of the load line. Hence if the value of E in Fig. 6-5*a* is too large or too small, there will be only one operating point. Similarly, if the load line is exactly tangent to the middle portion of the tunnel diode v_1-i_1 curve, there will be two operating points. In practice, of course, only one operating point is observable at any one time. Which of these *possible* operating points is actually attained at a given time is decided by the parasitic elements which are invariably present. The methods for obtaining this important information are presented in Part 3 of this book.

It is important to recognize that the load-line method is not restricted to a network containing only two resistors, a most unlikely if not trivial situation. On the contrary, the use-

Fig. 6-5. A simple tunnel-diode switching circuit.



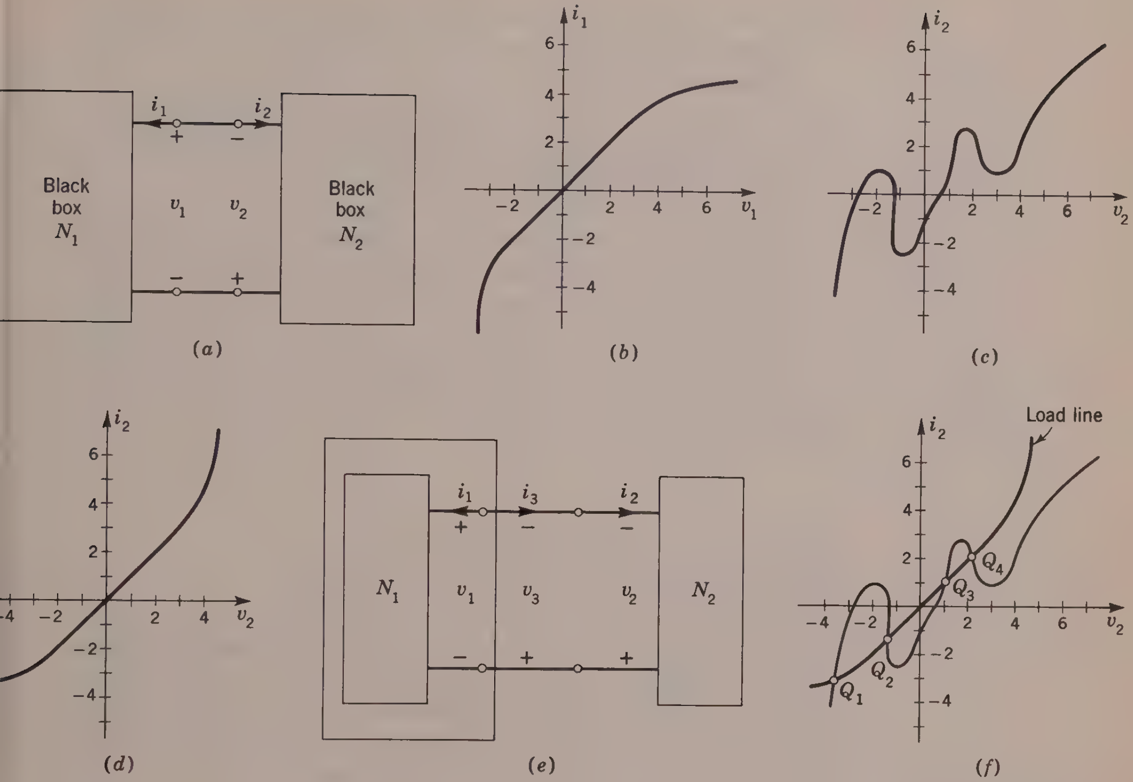


Fig. 6-6. The graphical procedure for determining the operating point of two resistive black boxes connected in the form of basic configuration 1 by the load-line method.

fulness of the load-line method lies in the fact that there are many practical networks whose elements can be combined into two separate black boxes in the form of basic configuration 1.

2. Consider two black boxes N_1 and N_2 connected as shown in Fig. 6-6*a*. Each black box may contain an arbitrary interconnection of other resistive network elements such as zener diodes, transistors, and gyrators. As far as we are concerned, however, only the external DP plots of N_1 and N_2 are necessary for determining the operating points. Assume for the time being that these have been found and are given in Fig. 6-6*b* and *c*, respectively.[1] Notice that although this network is in the form of basic configuration 1, the references of the current and voltage variables (for which the DP plots are based) are not identical with those shown in Fig. 6-3*a*. Since the v_1-i_1 curve is simpler, the black box N_1 is chosen to be the load. The laws of interconnection are $i_1 = -i_2$ and $v_1 = -v_2$. Hence, the load line is obtained by reflecting the v_1-i_1 curve, first with respect

---

[1] For reasons that will soon be obvious, it may be more convenient to choose a particular set of reference voltage polarity and reference current direction in the process of determining the DP plot of each black box. Hence, we must always watch for the references associated with a given DP plot.

to the v_1 axis, and then with the i_1 axis. The net effect is equivalent to a 180° rotation with respect to the origin as shown in Fig. 6-6*d*.

The load line in Fig. 6-6*d* could also be obtained by redefining a new set of terminal variables i_3 and v_3 for the black box N_1 as shown in Fig. 6-6*e*. Since $i_3 = -i_1$ and $v_3 = -v_1$, the new v_3-i_3 curve coincides with the load line in Fig. 6-6*d*. In either case, the intersections between the load line, or the v_3-i_3 curve, with the v_2-i_2 curve give the four operating points Q_1, Q_2, Q_3, and Q_4 shown in Fig. 6-6*f*. Using the voltage and current values at each operating point, it is possible to determine the voltage and current solutions of the network elements inside N_1 and N_2. The procedure for doing this will be discussed in the next chapter.

6-2-2 BASIC NETWORK CONFIGURATION 2, INTERCONNECTION BETWEEN 2 TWO-TERMINAL RESISTORS AND A THREE-TERMINAL RESISTOR

Another very common network configuration of practical interest is shown in Fig. 6-7. To be specific, let the three-terminal resistor be characterized by the set of characteristic curves shown in Fig. 6-8*a* and *b*, and let the nonlinear resistors R_3 and R_4 be characterized by the curves shown in Fig. 6-8*c* and *d*, respectively. The equations of motion for this network are as follows:

Equations from the laws of elements

Three-terminal resistor:

$$\begin{cases} i_1 = g_1(v_1,i_2) & (6\text{-}3a) \\ i_2 = g_2(v_2,v_1) & (6\text{-}3b) \end{cases}$$

Resistor R_3: $\quad i_3 = g_3(v_3) \qquad$ (6-3*c*)

Resistor R_4: $\quad i_4 = g_4(v_4) \qquad$ (6-3*d*)

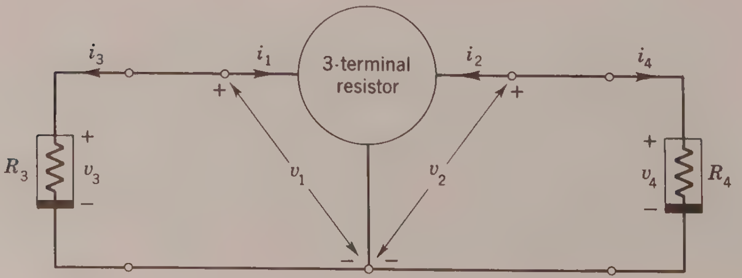


Fig. 6-7. The basic network configuration 2 consists of a nonlinear resistor connected across the input and the output ports of a three-terminal resistor.

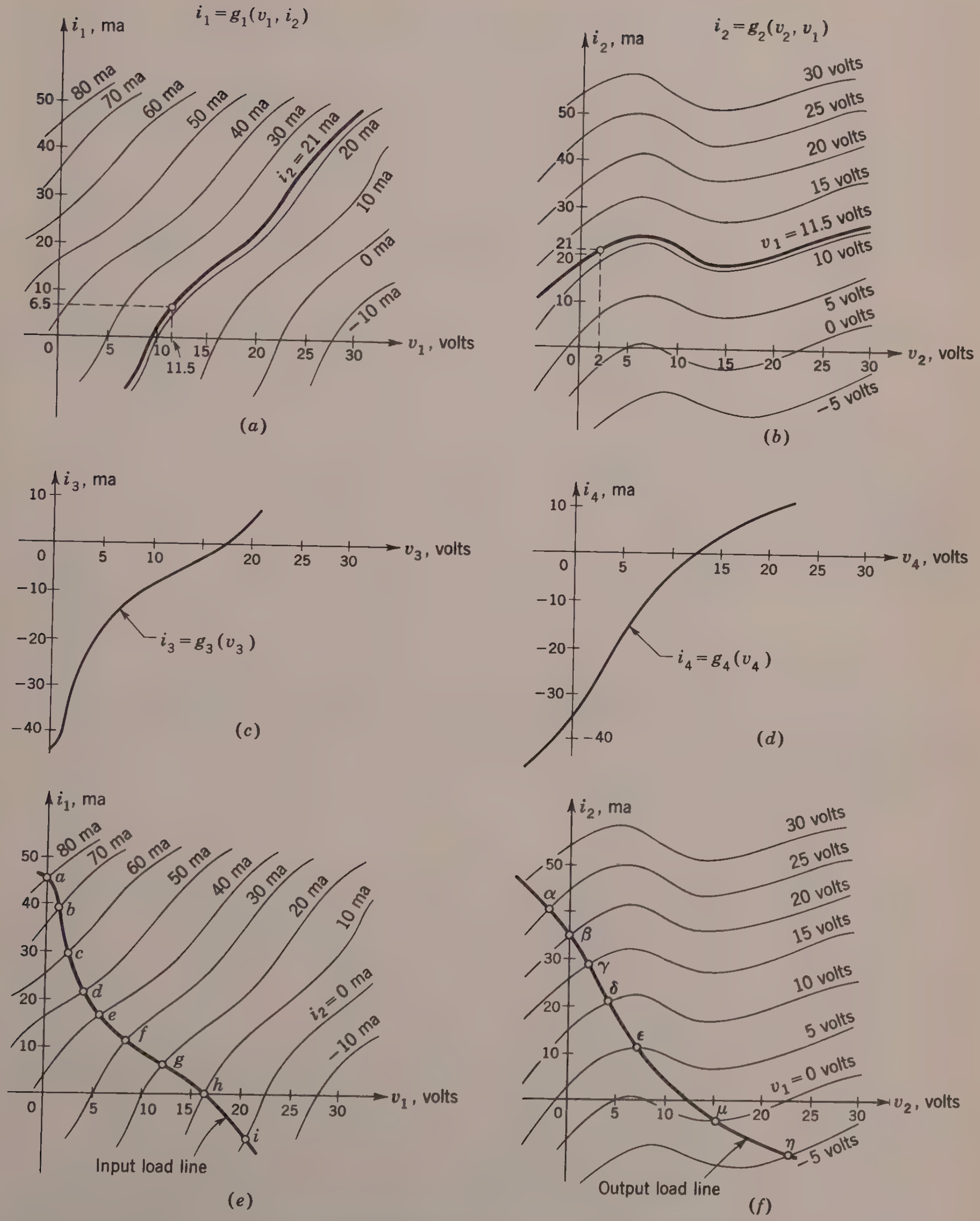


Fig. 6-8 The graphical procedure for determining the operating point of a resistive nonlinear network in basic configuration 2 by the double load-line method.

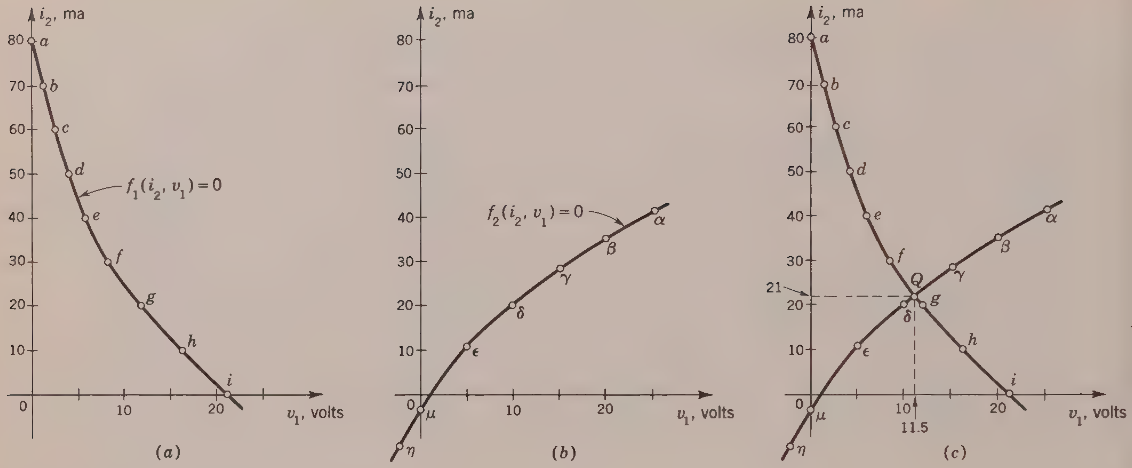


Fig. 6-9. When the input load line and the output load line are transferred into the same v_1-i_2 plane, their intersection gives the location of the operating point.

Equations from the laws of interconnection

$$v_3 = v_1 \tag{6-4a}$$

$$v_4 = v_2 \tag{6-4b}$$

$$i_3 = -i_1 \tag{6-4c}$$

$$i_4 = -i_2 \tag{6-4d}$$

To eliminate the variables v_3, v_4, i_3, and i_4, we substitute Eqs. (6-4*a*), (*b*), (*c*), and (*d*) into Eqs. (6-3*c*) and (*d*) and obtain

$$i_1 = -g_3(v_1) \tag{6-5a}$$

$$i_2 = -g_4(v_2) \tag{6-5b}$$

Observe that Eq. (6-5*a*) represents a curve obtained by reflecting the v_3-i_3 curve with respect to the v_3 axis, and then relabeling the variables v_3 and i_3 by v_1 and i_1. This curve is drawn on top of the input-characteristic curves of the three-terminal resistor as shown in Fig. 6-8*e* and is called the *input load line*. Similarly, Eq. (6-5*b*) represents a curve obtained by reflecting the v_4-i_4 curve with respect to the v_4 axis and then relabeling the variables v_4 and i_4 by v_2 and i_2, respectively. This curve is drawn on top of the output-characteristic curves of the three-terminal resistor as shown in Fig. 6-8*f* and is called the *output load line.*

Equations (6-5*a*) and (*b*) constitute two equations in four unknowns. It is necessary, therefore, to eliminate two more variables before the operating point can be located. This can be achieved graphically by observing that each point of intersection P between the input load line and the input-characteristic curves in Fig. 6-8*e* specifies the values of v_1, i_1, and i_2 at P. Similarly, each point of intersection P between the output load line and the output-characteristic curves in Fig. 6-8*f* specifies the values of v_2, i_2, and v_1 at P. The key for eliminating the variables consists in identifying the variables which are common in both sets of curves. In this case, they are v_1 and i_2. Hence, the input load line in Fig. 6-8*e* can now be transferred into the v_1-i_2 plane as shown in Fig. 6-9*a*. Similarly, the output load line can be transferred into the v_1-i_2 plane as shown in Fig. 6-9*b*. The point Q where these two curves intersect is the operating point. This is shown in Fig. 6-9*c* where the coordinates for Q are given by $v_1 = 11.5$ volts and $i_2 = 21$ ma. We can now locate this point in the input-characteristic curves of Fig. 6-8*a* and read the corresponding value of i_1, namely, $i_1 = 6.5$ ma. Similarly, the value of $v_2 = 2.0$ volts can be obtained by locating the point Q on the output-characteristic curves of Fig. 6-8*b*.

A review of the above systematic procedure would suggest some shortcuts. For example, instead of plotting the input and output load lines separately as in Fig. 6-9, we could have simply transferred the coordinates v_1 and i_2 of the output load line from Fig. 6-8*f* onto the input-characteristic curves of Fig. 6-8*e*. The point of intersection between these two load lines as shown in Fig. 6-10*a* gives exactly the same answers. Alternately, we could transfer the input load line from Fig. 6-8*e* onto the output-characteristic curves of Fig. 6-8*f* as shown in Fig. 6-10*b*. Again, the same answers are obtained, as they should be. In practice, some further shortcuts may be possible. For example, it is not

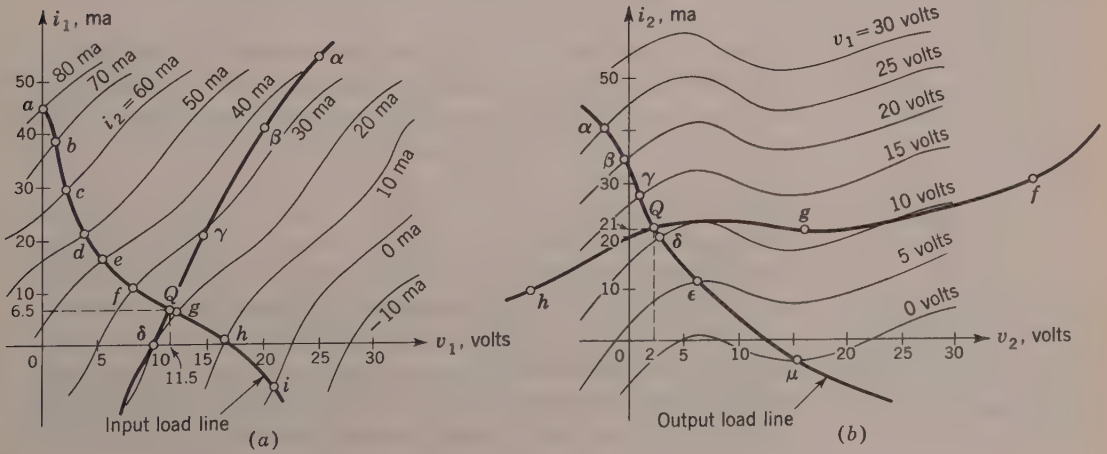


Fig. 6-10. Two alternate shortcuts for locating the operating point of a resistive nonlinear network in basic configuration 2.

really necessary to transfer the entire curve into Fig. 6-10*a* or Fig. 6-10*b*. After plotting a few trial points, we could usually zero in by plotting only points which are near the other load line. In fact, the input-characteristic curves of many three-terminal resistors are usually so crowded near each other that only a few trial-and-error steps are necessary to locate the operating point.

We shall refer to the above graphical procedures as the *double load-line method.* In addition to the many practical networks already in this configuration, there are many more complicated networks which can be reduced to the same configuration by an appropriate grouping of elements into black boxes.

EXAMPLE

Consider the transistor circuit shown in Fig. 6-11*a*, where each of the two black boxes contains a resistor-battery combination. The input- and output-characteristic curves of the *n-p-n* transistor are shown in Fig. 6-11*b* and *c*, respectively.[1] The equations for the input and output load lines are readily obtained from Fig. 6-11*a*; thus

$$V_{BE} = -R_1 I_B + E_1 \tag{6-6a}$$

$$V_{CE} = -R_2 I_C + E_2 \tag{6-6b}$$

These load lines are drawn on top of the input- and output-characteristic curves in Fig. 6-11*b* and *c*, respectively.[2] Since the input characteristics consist of only one curve, the intersection between this curve and the input load line immediately yields two of the coordinates of the operating point, namely, $V_{BE} = V_{BE_Q}$ and $I_B = I_{B_Q}$. Since I_B is the parameter for the output-characteristic curves, the value $I = I_{B_Q}$ singles out one curve (shown heavy) from the rest as shown in Fig. 6-11*c*. The intersection between this I_C-vs.-V_{CE} curve (corresponding to $I_B = I_{B_Q}$) and the output load line gives the remaining coordinates of the operating point Q, namely, $I_C = I_{C_Q}$ and $V_{CE} = V_{CE_Q}$.

This example clearly demonstrates the computational advantages resulting from approximating the input-characteristic curves by a single curve. However, it must be emphasized that this shortcut is valid only if the input-characteristic curves of the three-terminal resistor are closely spaced. Fortunately, two of the most useful devices, transistors and FET, do satisfy this property. One might even be tempted to assert that it is precisely this property

---

[1] Notice that we have represented the input-characteristic curves by a single curve rather than by a family of closely packed curves as shown in Table 3-1. Approximating these curves by one curve is accurate enough for most practical purposes.

[2] These types of load lines are commonly encountered in practice. It is convenient, therefore, to remember that the voltage intercept in each case is equal to the battery voltage, and the slope in each case is equal to the negative reciprocal of the respective load resistance.

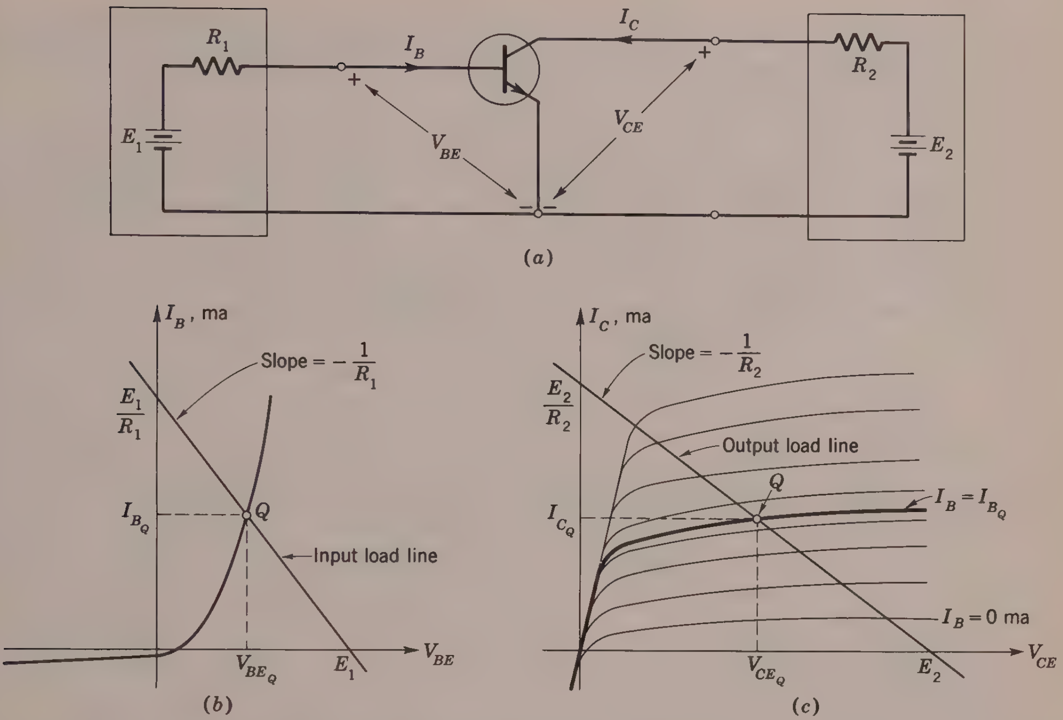


Fig. 6-11. The double load-line method for locating the operating point of a transistor circuit terminated by two pairs of resistor-battery combinations.

that makes these devices useful because it makes the design task simpler. This may be true to a certain extent; but one must be more objective, for after all, there are other useful devices such as the unijunction transistor (see Table 3-1) which do not fall into this category. It is important, therefore, to learn the general procedure since it is valid for all cases.

The double load-line method is applicable to any network which can be redrawn into the form of basic configuration 2, provided the DP plot of the two external loads can be found. For example, consider the network shown in Fig. 6-12*a*. At first sight, this seems to be a formidable problem because of the large number of elements present. To be sure, the equations of motion of this network are certainly going to be quite complicated. However, careful inspection suggests that we can redraw this network in the form of basic configuration 2 by grouping elements into the two black boxes N_1 and N_2 shown in Fig. 6-12*b*. The double load-line method is therefore applicable here, provided we can find the DP plots of N_1 and N_2.[1] The necessity for determining DP plots of resistive nonlinear networks is one important motivation for us to learn the methods for accomplishing this task.

---

[1] The first fundamental theorem of resistive nonlinear networks in Chap. 5 guarantees that such DP plots exist.

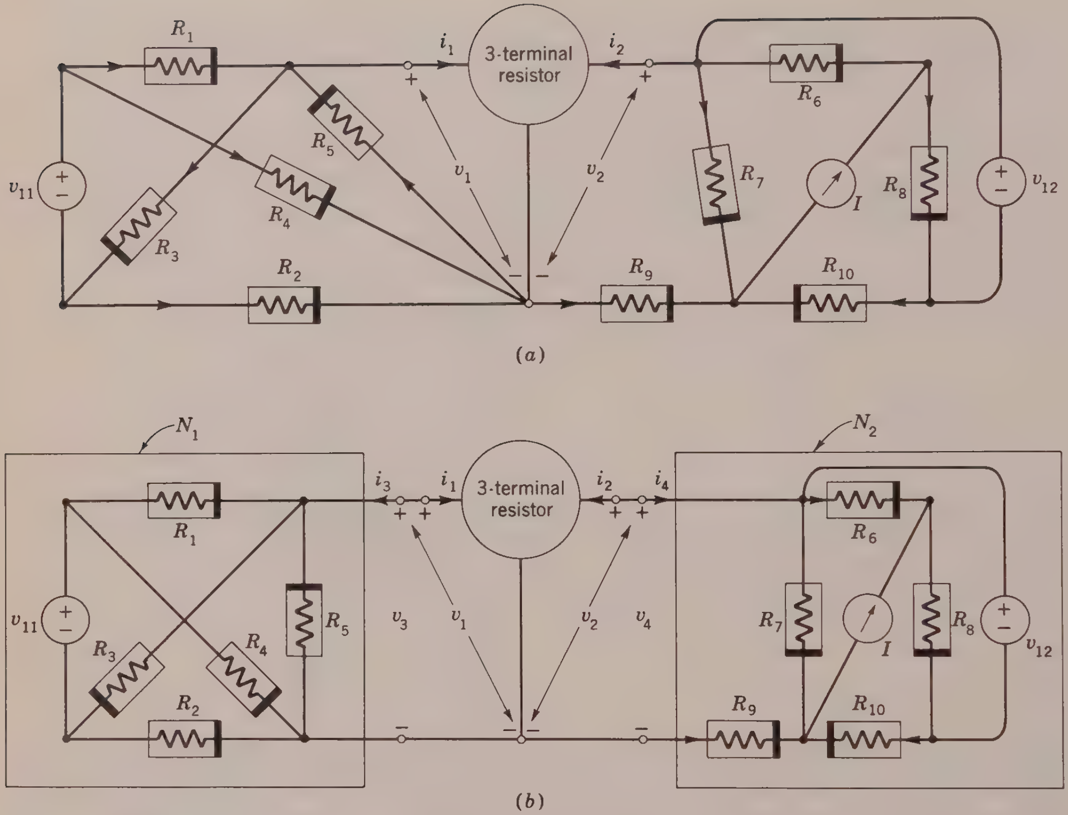


Fig. 6-12. Example of a complicated network which can be redrawn into the form of basic configuration 2.

Exercise 1: Consider the tunnel diode network shown in Fig. 6-5. (*a*) If $R = 500\ \Omega$, find the range of values of E for which there will be one, two, or three operating points. (*b*) If $E = 0.45$ volt, find the range of values of R for which there will be one, two, or three operating points.

Exercise 2: Give an alternate interpretation of the double load-line method by redefining the terminal variables of resistors R_3 and R_4 in Fig. 6-7.

Exercise 3: Sketch the graphical construction for determining the operating point of the network shown in Fig. 6-11, but with resistors R_1 and R_2 replaced by two identical varistors.

6-3 GRAPHICAL DETERMINATION OF DP PLOTS OF SERIES-PARALLEL NETWORKS

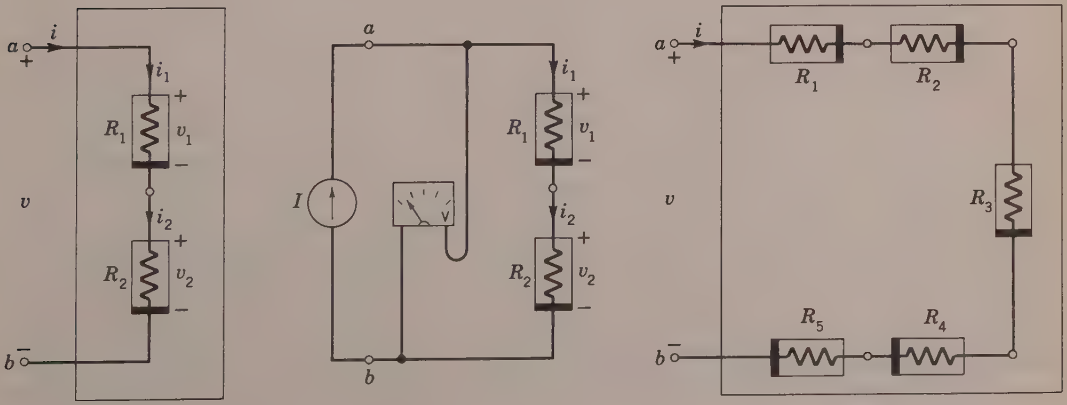
Using graphical methods, we can readily obtain the DP plot across the driving-point terminals of any black box containing only series-parallel interconnection of two-terminal resistors and dc sources. The procedure consists of a repeated application of two basic graphical techniques, namely, the *series-combination technique* and the *parallel-combination technique.*

6-3-1 THE SERIES-COMBINATION TECHNIQUE

Consider first the simplest case of two nonlinear resistors connected in series as shown in Fig. 6-13*a*. Since the negative terminal of R_1 is connected to the positive terminal of R_2, we shall call this the *back-to-front series connection.* One method for determining the DP plot of this network consists of applying an adjustable current source I across terminals *a-b* and measuring the corresponding voltage. Since $I = i_1 = i_2$ and $v = v_1 + v_2$, the DP plot can be obtained by first aligning the v_1 axis with the v_2 axis, and then for each value of current $i_1 = i_2 = I$ graphically adding the value of v_1 from the v_1-i_1 curve, and the value of v_2 from the v_2-i_2 curve. This step can be carried out easily by drawing horizontal guidelines. If there are more than two resistors, the same procedure can obviously be applied, provided all resistors are connected in the back-to-front configuration shown in Fig. 6-13*c*.

EXAMPLES

1. Suppose the two resistors of Fig. 6-13*a* are linear with resistances R_1 and R_2. The *v-i* curves corresponding to R_1 and R_2 are, therefore, straight lines as shown in Fig. 6-14*a* and *b*. The DP plot of the two linear resistors in series is readily obtained by the above graphical procedure and is shown in Fig. 6-14*c*. Observe that since two points determine a straight line, we



(*a*) (*b*) (*c*)

Fig. 6-13. A string of back-to-front series-connected nonlinear resistors.

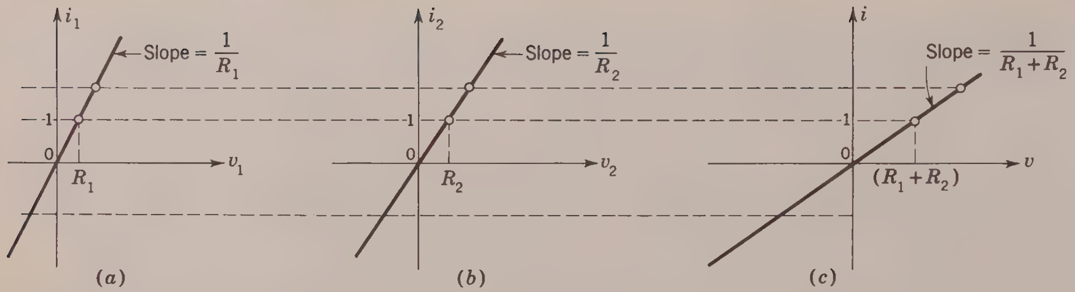


Fig. 6-14. The graphical procedure for determining the DP plot of two linear resistors in series.

only have to apply this graphical procedure to two values of current. One convenient value of current to choose for this example is obviously $i_1 = i_2 = 0$.

2. Consider next the case where resistors R_1 and R_2 are characterized by the v_1-i_1 and v_2-i_2 curves shown in Fig. 6-15*a* and *b*, respectively. The DP plot obtained by the graphical procedure is shown in Fig. 6-15*c*. Although only a few horizontal guidelines

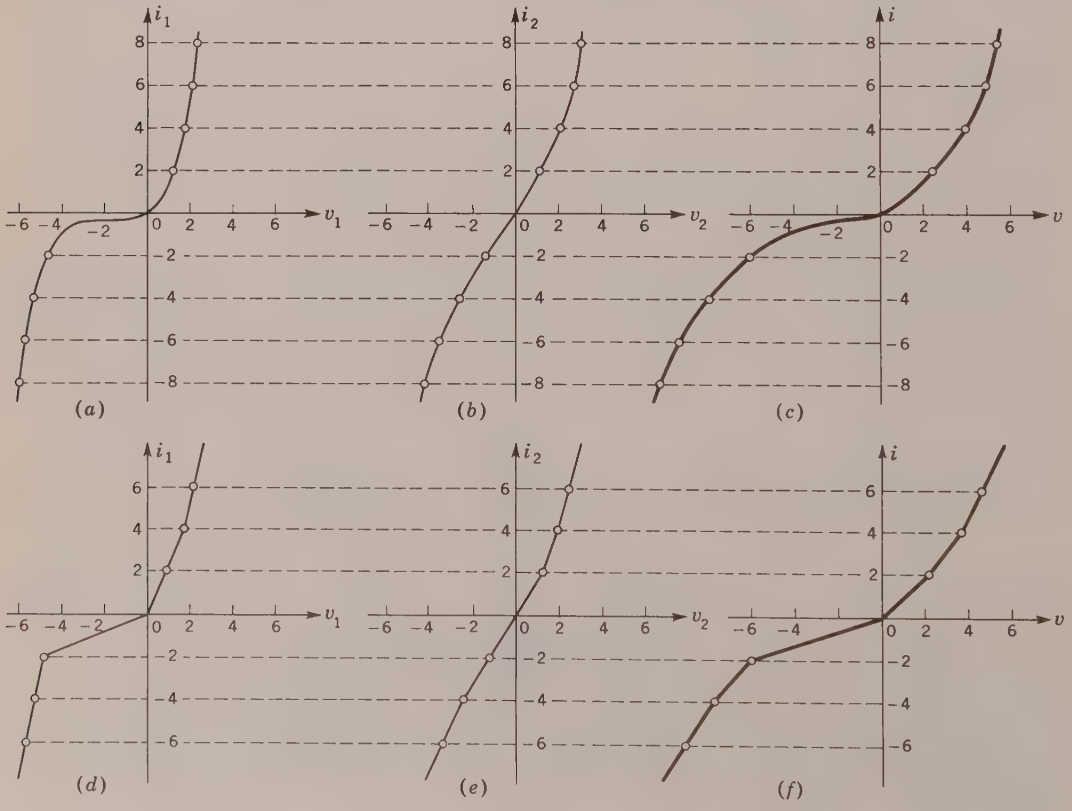


Fig. 6-15. The DP plot of two nonlinear resistors in a back-to-front series connection can be obtained by adding the voltage of each curve along each horizontal guideline.

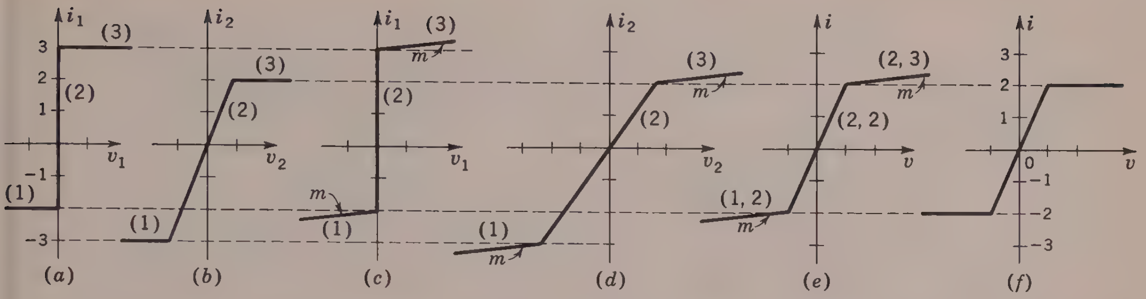


Fig. 6-16. The DP plot corresponding to v-i curves with horizontal segments can best be determined by a limiting process.

are shown, it is clear that more points must be found in order to obtain an accurate solution. This could be a rather time-consuming process. In most practical cases, we can simplify the task by first approximating the v-i curves with piecewise-linear segments as shown in Fig. 6-15*d* and *e*. From Example 1 we know two things: (1) The DP plot will also be piecewise-linear. (2) The horizontal guidelines need be drawn only through the breakpoints on each curve and one convenient point on each end segment. The resulting DP plot for this case is shown in Fig. 6-15*f*.

The above piecewise-linear procedure is relatively simple to carry out for most cases except when the v-i curves contain one or more horizontal segments. The confusions that sometimes resulted from the presence of horizontal segments can be resolved by a number of ways. One way is to interpret the horizontal segment as an open circuit and make use of the fact that any element in series with an open circuit results in another open circuit. The corresponding segment of the DP plot must therefore be horizontal. Another and perhaps more satisfactory way to resolve the above confusion is to replace each horizontal segment by a segment with a small but finite slope. The DP plot can then be determined by a limiting process. For example, suppose the v_1-i_1 and v_2-i_2 curves of R_1 and R_2 are shown in Fig. 6-16*a* and *b*, respectively. Since the end segments of both v-i curves are horizontal, let us replace them by the near-horizontal segments with a small positive slope m, as shown in Fig. 6-16*c* and *d*, respectively. If we label each segment combination consisting of segment n_1 of R_1 and n_2 of R_2 by (n_1, n_2), then segment (1,2) in Fig. 6-16*e* is obtained by adding the abscissas of segment 1 of R_1, segment 2 of R_2, etc.[1] If we gradually reduce the slope m to zero in the limit, we obtain the DP plot shown in Fig. 6-16*f*. In this case, segments (1,1) and (3,3) are located at $\pm\infty$, respectively.

[1] Notice that segments (1,1) and (3,3) are not shown because they are located off the page.

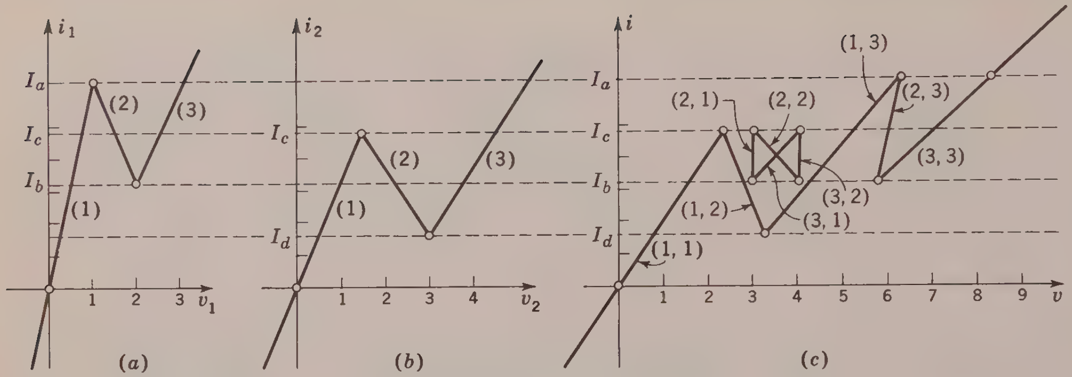


Fig. 6-17. A nine-segment multivalued DP plot is obtained by combining two appropriate tunnel diodes in series.

3. Let us now consider the case where the v-i curves are not monotonic. In particular, suppose R_1 and R_2 are tunnel diodes with the v_1-i_1 and v_2-i_2 curves shown in Fig. 6-17*a* and *b*, respectively. Since some horizontal guidelines would intersect the v_1-i_1 or the v_2-i_2 curves at more than one point, we must be more careful in carrying out the graphical procedure. If we recall that each point on the DP plot must correspond to the operating point of the network, then it is clear that if there are more than one operating points at a given current $i = I$, the DP plot must have a corresponding number of points at $i = I$. Hence, the DP plot will be *multivalued.* In order that we do not miss any operating points while carrying out the graphical procedure, it is important that we consider one segment of the v_1-i_1 curve at a time and combine this segment with *each* of the segments of the v_2-i_2 curve. For example, let us pick segment 2 of the v_1-i_1 curve first. Observe that this segment is defined only for $I_b \leq i_1 \leq I_a$. To combine this segment with segment 1 of the v_2-i_2 curve, which is defined only for $-\infty \leq i_2 \leq I_c$, we must first determine a range of current for which both segments are defined. In this case, the common range is $I_b \leq i \leq I_c$. Any current outside this range will not intersect one of the two segments, and hence no operating point is possible. The next step is to apply the above graphical procedure and obtain segment (2,1) as shown in Fig. 6-17*c*. Repeating the same procedure for segment 2 of R_1 and segment 2 of R_2, we obtain segment (2,2). Similarly, combining segment 2 of R_1 and segment 3 of R_2 leads to segment (2,3). As soon as all combinations with segment 2 of R_1 are exhausted, we pick another segment and repeat the procedure all over again. For the present example, the v_1-i_1 and the v_2-i_2 curves have three seg-

ments each. Accordingly, there is a total of nine combinations. The reader should verify that the DP plot shown in Fig. 6-17*c* is indeed obtained by this procedure.

Before going to the next section, it must be emphasized that the above graphical procedure is valid under only two assumptions: (1) The resistors are connected in the back-to-front manner shown in Fig. 6-13. (2) The usual reference voltage polarity and current direction are chosen for each element. Since the same procedure is applicable to the case where the resistors are replaced by black boxes (which may contain other elements inside), the above two assumptions must be carefully checked. If the reference polarity and direction do not satisfy these conditions, they must first be redefined properly before the graphical procedure is carried out.

Exercise 1: Consider the network shown in Fig. 5-12*b* consisting of two junction diodes connected back to back in series. (*a*) Sketch the DP plot of this network. (*b*) If the two diodes are replaced by ideal diodes, find the DP plot. HINT: Redefine the polarity and direction of one of the two diodes.

Exercise 2: Show that the DP plot of any resistor with a v-i curve Γ in series with a battery of terminal voltage E is simply a horizontal translation of Γ by E volts along the v axis.

Exercise 3: (*a*) Show that the DP plot segment (s_1,s_2) corresponding to segments s_1 of R_1 and s_2 of R_2 does not exist if the range of definition for segment s_1 and the range of definition for segment s_2 do not overlap. (*b*) Show that if the v_1-i_1 curve has n_1 segments and the v_2-i_2 curve has n_2 segments, then the DP plot will have n segments, where $n \leq n_1 n_2$. Why is the inequality sign needed?

6-3-2 THE PARALLEL-COMBINATION TECHNIQUE

Consider next the back-to-back parallel connection of two nonlinear resistors as shown in Fig. 6-18*a*. The reader will recognize this to be the dual of the series circuit presented earlier. Indeed, if we construct the dual graph of the series circuit of Fig. 6-13*b* as shown in Fig. 6-18*b*, we obtain the parallel network shown in Fig. 6-18*a*. Hence, the procedure for determining the DP plot is the dual of the procedure for the series case. In order to avoid interchanging the current and the voltage axes, we could carry out the graphical procedure by aligning the current axes and by drawing vertical guidelines. For example, corresponding to the v_1-i_1 and the v_2-i_2 curves shown in Fig. 6-18*c* and *d*, we obtain the DP plot shown in Fig. 6-18*e*.

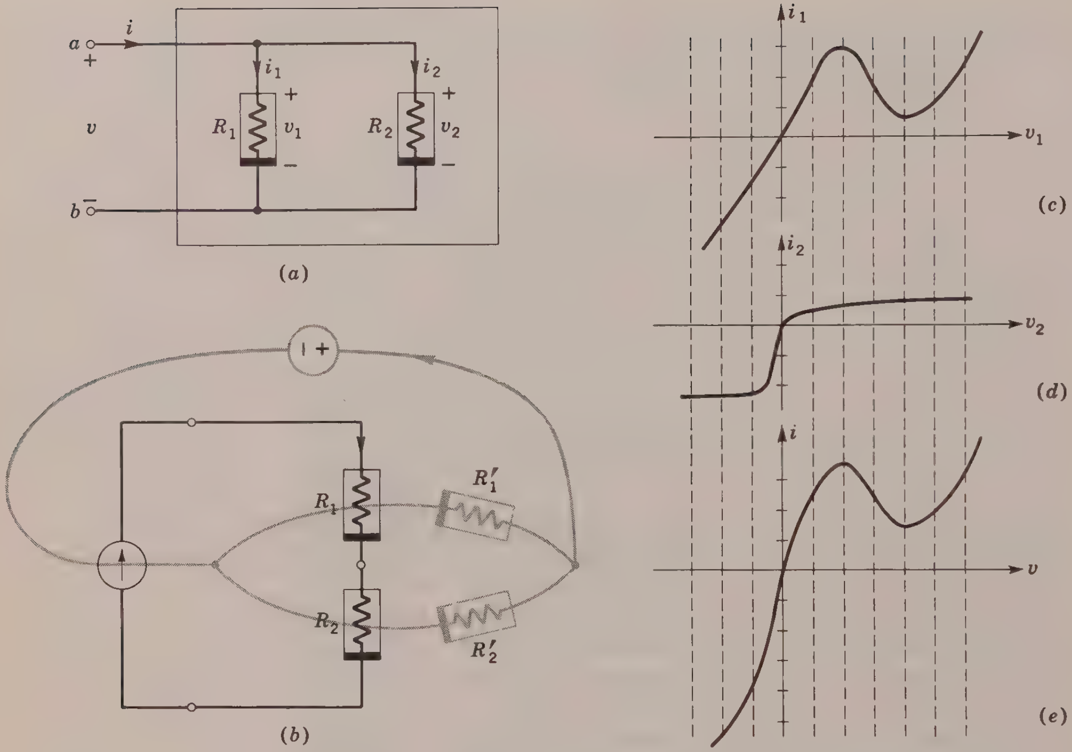


Fig. 6-18. A back-to-back parallel network and the graphical procedure for obtaining its DP plot.

Exercise 1: Consider two junction diodes connected in parallel (with opposite polarities). (*a*) Find the DP plot. (*b*) Replace the diodes by ideal diodes and obtain the DP plot.

Exercise 2: Discuss the procedure for finding the DP plot of a parallel network of nonmonotonic resistors.

Exercise 3: Show that the DP plot of any resistor with v-i curve Γ in parallel with a current source with terminal current I is simply a vertical translation of Γ by I amperes along the i axis.

6-3-3 COMBINATION OF SERIES-PARALLEL TECHNIQUES

By a repeated application of the series-combination and the parallel-combination techniques, the DP plot across any series-parallel network can be obtained. The basic idea is to replace all resistors in series or in parallel by a new resistor whose v-i curve is the DP plot of the two-terminal black box containing the resistors. Each application of this reduction technique will reduce the number of resistors in series or in parallel. The reduction technique is applied repeatedly until the network reduces to one resistor.[1]

As an illustration of the reduction technique, consider the network N shown in Fig. 6-19*a*. To emphasize that N is a series-parallel network, it is redrawn as shown in Fig. 6-19*b*. We can now apply the series-combination technique by replacing resistors R_1 and R_2 by R_{18}, and resistors R_{16} and R_{17} by R_{22}, as shown in Fig. 6-19*c*. Similarly, we apply the parallel-combination technique by replacing resistors R_5 and R_6 by R_{19}, resistors R_7 and R_8 by R_{20}, and resistors R_{11} and R_{12} and the current source by R_{21}. But now with the new resistors, there are new groups of series-parallel resistors. These new groups can be reduced further. We can reduce the network in Fig. 6-19*c* by replacing parallel resistors R_3 and R_{18} by R_{23}, series resistors R_9 and R_{20} by R_{24}, and parallel resistors R_{15} and R_{22} by R_{25}, as shown in Fig. 6-19*d*. Of course, this circuit can be reduced still further. For example, resistor R_4, R_{23} can be replaced by resistor R_{26}, as shown in Fig. 6-19*e*. Resistor R_{26} can then be combined with R_{19} to form R_{27} in Fig. 6-19*f*. By this time, we can see that by applying the reduction procedure a few more times to the chain of black boxes shown in Fig. 6-19*f*, the network would eventually reduce to only one resistor whose DP plot is the desired solution.

6-4 SOME PRACTICAL APPLICATIONS OF DP PLOT AND OPERATING-POINT CONCEPTS

The techniques presented so far can be applied to a surprisingly large class of practical electronic circuits. While it is true that most practical circuits are more complex than those discussed earlier, many of these can be reduced to simpler circuits by combining appropriate groups of elements into black boxes. Once the DP plots of these black boxes are found, the problem becomes a lot simpler. One of the objectives of this book is to bring forth the power of this black-box approach. Let us consider some concrete examples.

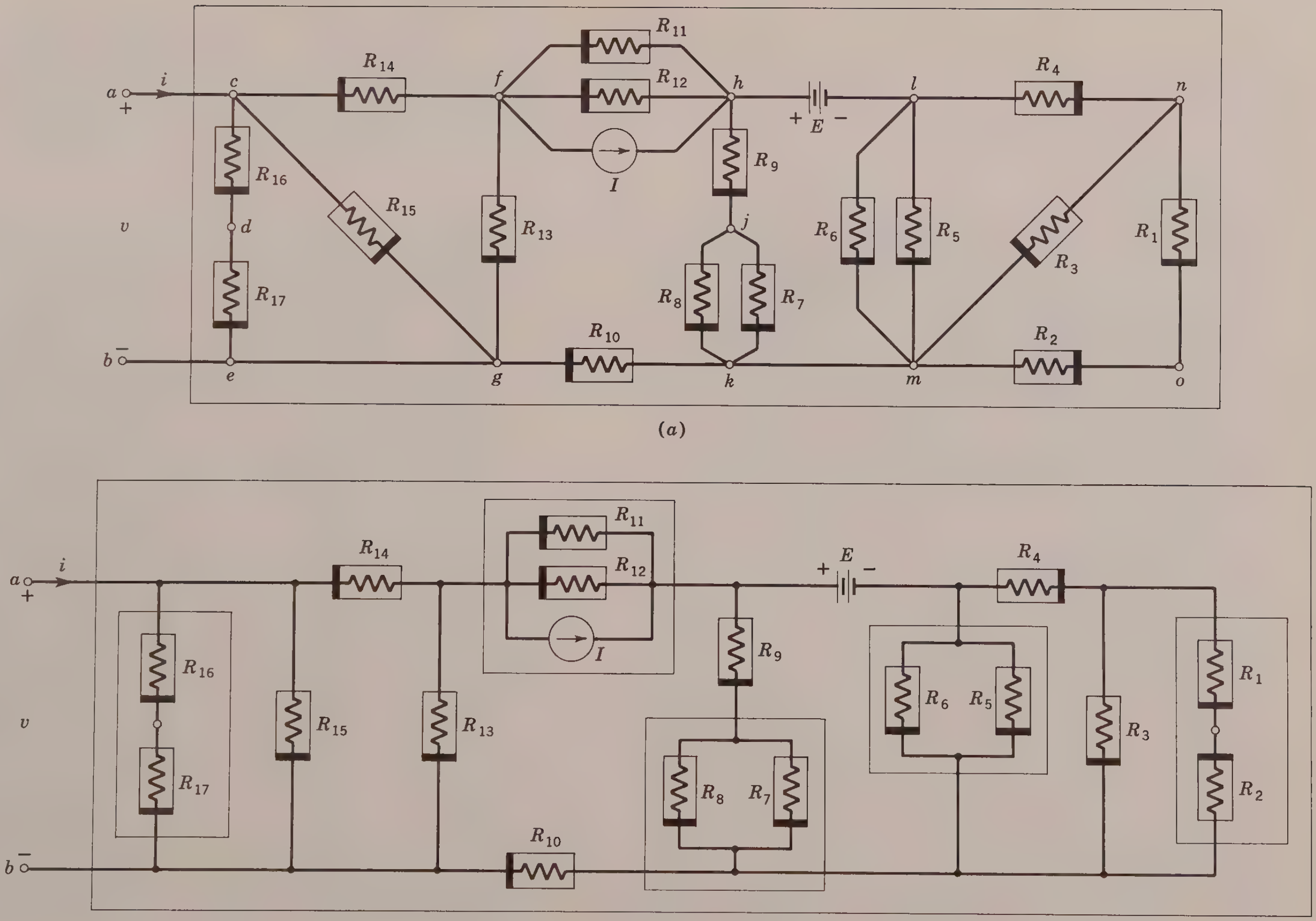
6-4-1 CONCAVE AND CONVEX RESISTORS

The black box shown in Fig. 6-20*a* is made up of a junction diode, a zener diode, and a linear resistor in series. It is called a *concave resistor.*[2] The v_1-i_1 curve of the junction diode and the v_2-i_2 curve of the zener diode are shown by the dotted curves in Fig. 6-20*b* and *c*, respectively. For simplicity, we shall approximate

---

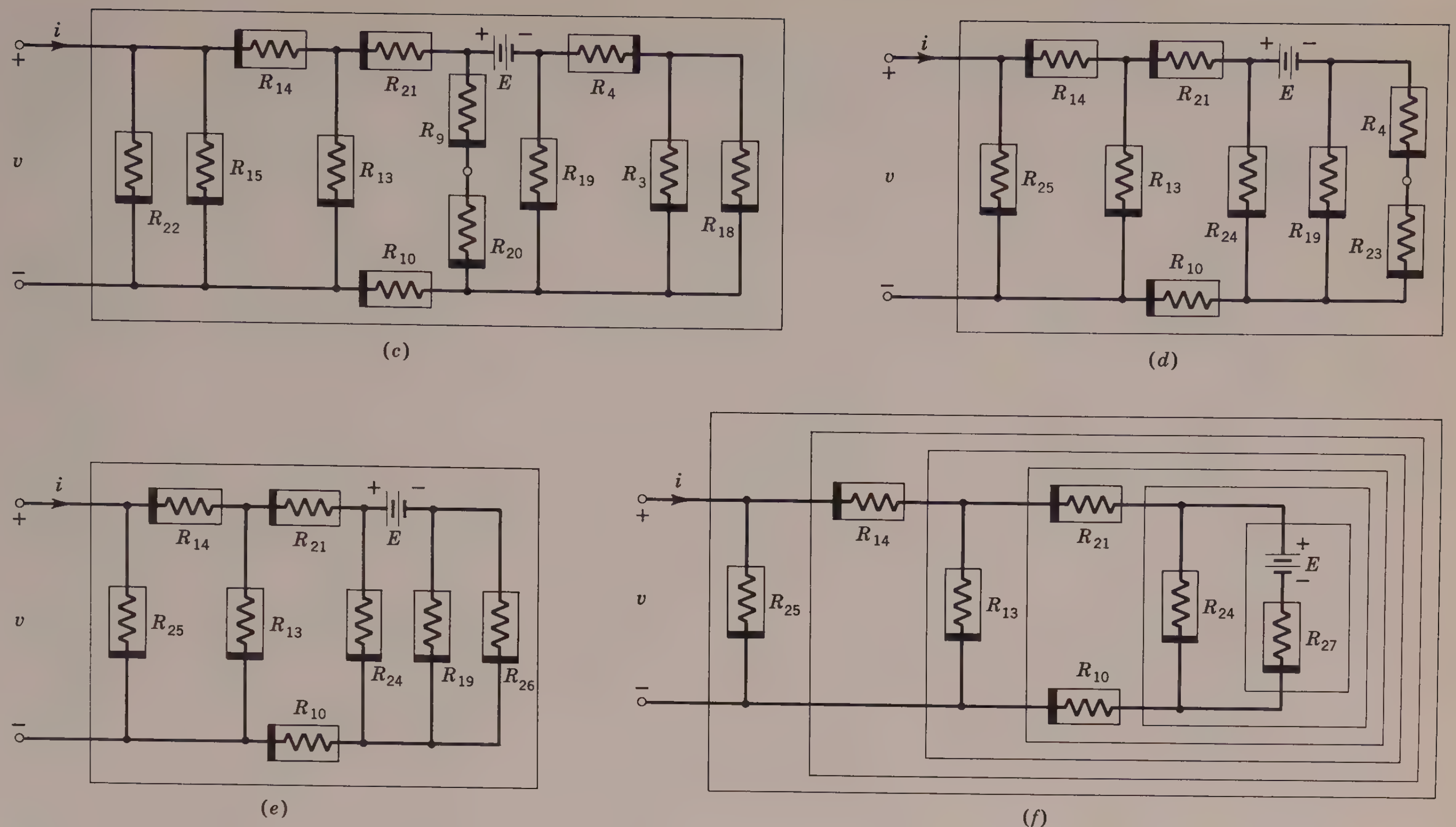
[1] A computer subroutine for implementing this algorithm can be easily written. A brief discussion of this subroutine is given in L. O. Chua, "A Computer Oriented Sophomore Course in Nonlinear Network Theory," *IEEE Trans. Circuit Theory,* Special Issue on Computer Applications in Education, September, 1969.

[2] A v-i curve is said to be concave if its slope di/dv is a nondecreasing function of v. It will be shown in Chap. 8 that concave resistors can be used as building blocks for synthesizing any prescribed concave v-i curve.


a
+
i
v
b
−
c
d
e
f
g
h
j
k
l
m
n
o
R_1
R_2
R_3
R_4
R_5
R_6
R_7
R_8
R_9
R_{10}
R_{11}
R_{12}
R_{13}
R_{14}
R_{15}
R_{16}
R_{17}
I
E
+
−

(a)

(b)



(c)

(d)

(e)

(f)

Fig. 6-19. The step-by-step reduction techniques for determining the DP plot of a series-parallel network.

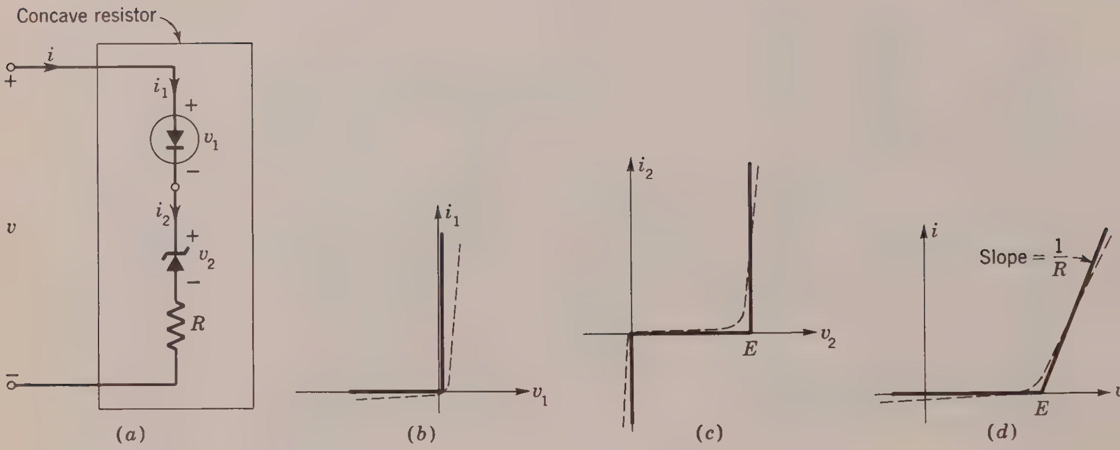


Fig. 6-20. By changing the value of the zener breakdown voltage E and the resistance R, the breakpoint and slope of the DP plot of the concave resistor can be easily adjusted to suit a particular application.

these curves by the piecewise-linear solid segments shown in the same figures. The DP plot of this black box is easily seen to be given by Fig. 6-20*d*, where the voltage E at the breakpoint is equal to the zener breakdown voltage (also known as the *avalanche voltage*), and the slope is equal to the reciprocal of the resistance R. This means that a concave resistor with any breakdown voltage E and slope $1/R$ can be easily synthesized in practice.[1] As such, concave resistors are indispensable building blocks for synthesizing arbitrary DP plots (see Chap. 8).

Consider next the black box shown in Fig. 6-21*a* consisting of a junction diode, a constant-current diode,[2] and a linear resistor. It is called a *convex resistor.*[3] The v_2-i_2 curve of the constant-current diode is shown by the dotted curve in Fig. 6-21*b*. For simplicity, we shall approximate this curve by the solid piecewise-linear segments shown on the same figure. Using this v_2-i_2 curve and an ideal diode curve for the junction diode, we obtain the DP plot shown in Fig. 6-21*c*. Observe that the breakpoint current and the slope of this DP plot can be adjusted by choosing an appropriate constant-current diode with the desired current I and an appropriate resistance R. Observe also that the v-i curve of the convex resistor is the dual of the concave resistor, and vice versa. Together, they constitute a fundamental set of building blocks for synthesis purposes (see Chap. 8).

6-4-2 ELEMENTS WITH A HORIZONTAL SEGMENT

A commonly used technique for obtaining a DP plot with a horizontal segment is to connect a linear resistor in parallel with a voltage-controlled resistor whose v-i curve contains a negative-resistance region.

---

[1] For more accurate results, a small correction factor representing the forward voltage drop (0.2 to 0.7 volt) of the junction diode must be added to the zener voltage E.

[2] This relatively new device is also known as a *currector* or as a *field-effect current-limiting diode.* It can be approximately realized in practice from an FET upon short-circuiting the gate and the source terminals. In this case, the solid horizontal segment in the left-half plane of Fig. 6-21*b* must be replaced by a vertical segment, thereby obviating the need for the junction diode in the convex resistor (see the footnote in Sec. 1-6-2).

[3] A v-i curve is said to be convex if its slope di/dv is a nonincreasing function of v. It will be shown in

A simple circuit consisting of a tunnel diode and a linear resistor in parallel is shown in Fig. 6-22*a*. The graphical construction for obtaining the DP plot is shown in Fig. 6-22*b*. Notice that if we choose the value of R equal to the magnitude of the tunnel-diode negative resistance, we would obtain the horizontal segment shown in this figure. In view of the low-voltage characteristics of the tunnel diode, this segment is horizontal only over a limited voltage range. To increase the range of the horizontal segment, a tunnel diode-battery combination can be added as shown in Fig. 6-22*c*. The v_a-i_a curve of the first tunnel diode is shown in Fig. 6-22*d*. Directly below it is the v_b-i_b DP plot of the tunnel diode-battery combination. The resulting v_1-i_1 DP plot is shown in Fig. 6-22*f*. Notice that the effect of adding the tunnel diode-battery combination is to compensate for each of the two negative-resistance segments so that they result in a single negative-resistance segment with a larger voltage range. The DP plot of the complete network is easily obtained and is shown in Fig. 6-22*g*. Notice that the operating range of the horizontal segment is almost three times that shown in Fig. 6-22*b*.

There are, of course, many other practical circuits using different devices that can achieve the same objective. However, the principles are quite similar. Which circuit to choose in any particular application depends on various practical considerations such as cost, operating range, linearity, and power rating.

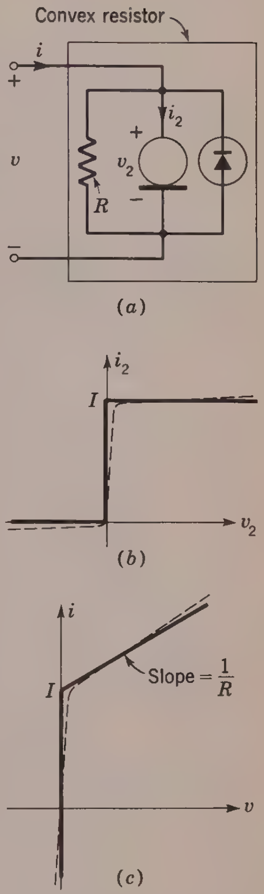


Fig. 6-21. By changing the values of I and R, the DP plot of a convex resistor can be adjusted to suit a particular application.

6-4-3 ELEMENTS WITH A VERTICAL SEGMENT

A practical technique for obtaining a DP plot with a vertical segment is to connect a linear resistor in series with a current-controlled resistor whose v-i curve contains a negative-resistance region. A widely used circuit consisting of an NTC thermistor is shown in Fig. 6-23*a*. Clearly, a vertical segment can be obtained by choosing the value of the resistance R to be equal to the magnitude of the negative resistance of the thermistor.

6-4-4 MULTITHRESHOLD ELEMENTS

In many practical applications such as switching circuits, a DP plot with several negative-resistance segments alternating with positive-resistance segments is required. A simple circuit for achieving this consists of a string of tunnel diodes in series as

Chap. 8 that convex resistors can be used as building blocks for synthesizing any prescribed convex v-i curve.

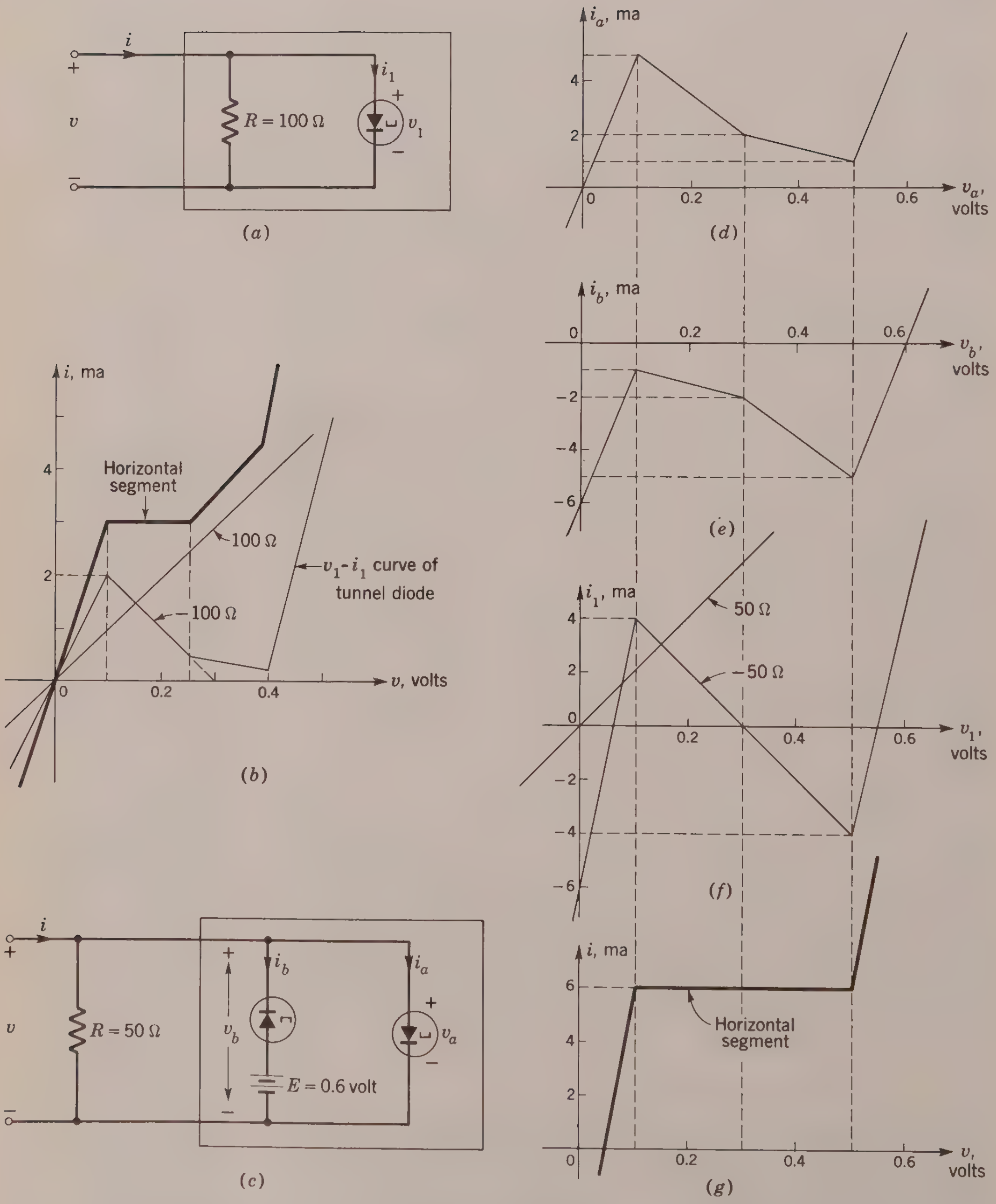


Fig. 6-22. Two practical tunnel-diode circuits for obtaining a DP plot with a horizontal segment.

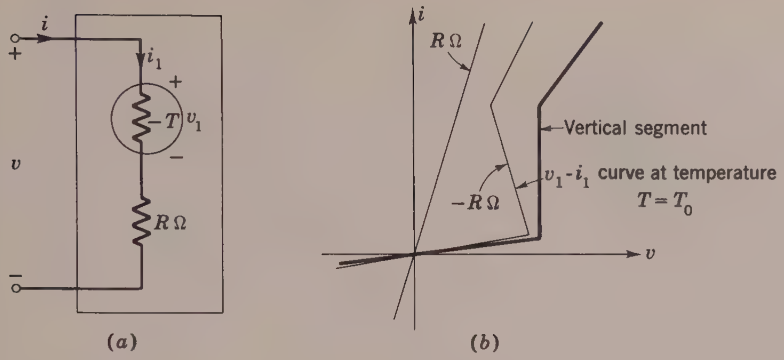


Fig. 6-23. The DP plot of an NTC thermistor in series with an appropriate linear resistor will contain a vertical segment at a given temperature $T = T_0$.

shown in Fig. 6-24*a*. Observe in Fig. 6-24*b* that the valley current of each tunnel diode was chosen greater than the peak current of the preceding tunnel diode.[1] This arrangement prevents the existence of additional segments, thereby making the graphical construction very easy. The resulting DP plot shown in Fig. 6-24*c* contains three negative segments alternating with four positive segments. Observe that if we excite the black box with a current source, and if we assume that the operating point is initially on segment 1, then as we increase the driving-point current i beyond $i = I_a$, the operating point will switch to segment 3. A further increase in the current to $i = I_c$ would switch the operating point from segment 3 to segment 5, etc. Obviously, by connecting more tunnel diodes in series, we can obtain any number (within the practical current range of available tunnel diodes, of course) of such alternating segments. If we monitor the voltage across the black box, then it is clear that each time the current i exceeds a *threshold* value (such as I_a, I_c, I_e, etc.), the voltage jumps to a higher value, thereby indicating a change in state. As such, this black box has found applications in multithreshold circuits. Observe that if we make the threshold currents I_a, I_c, and I_e differ only by a very small preset amount, we would obtain a very sensitive threshold device in the sense that a small change in current would be registered by the monitor, such as a voltmeter, as a relatively large change in voltage. This property is very useful in many control applications. For example, if several relays with different actuating voltages are connected across the black box, then each time a threshold current is exceeded, a relay will close, thereby actuating some external control mechanism.

---

[1] The currents at the minima and maxima of a tunnel diode v-i curve are often called the *valley* and the *peak currents*, respectively.

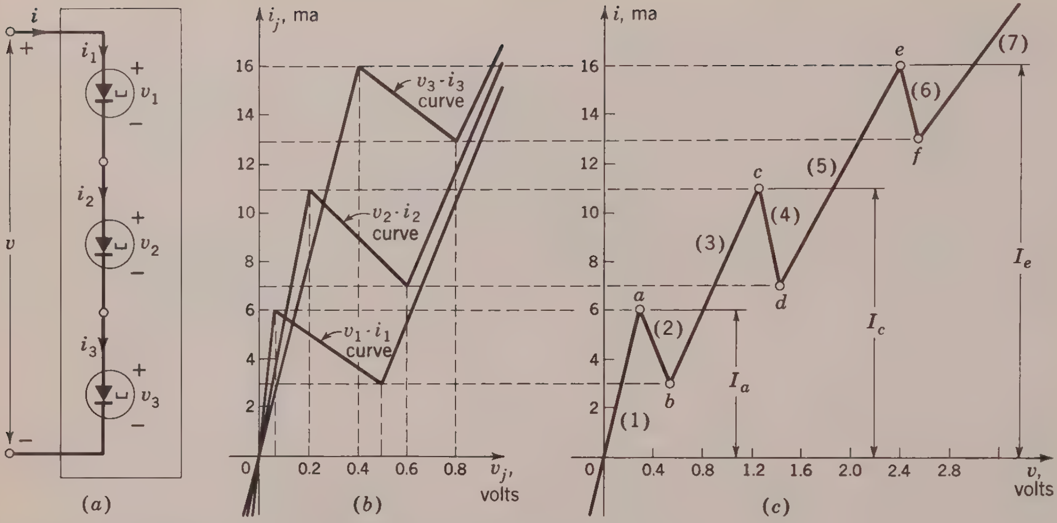


Fig. 6-24. A multithreshold DP plot can be realized by connecting a string of tunnel diodes in series.

6-4-5 AUTOMATIC SORTING CIRCUIT

One of the most common applications of the operating-point concept is the design of automatic sorting circuits for furnishing different preset output voltages in accordance with some "quality" of a product in the production line. A simple circuit of this nature is shown in Fig. 6-25*a*. This circuit is used in many beer bottling plants for sorting the liquid level of beer bottles on a production line. The basic idea stems from the fact that the intensity of a light beam is attenuated as it passes through a colored liquid such as beer. Hence, if we install a light source at the desired liquid level and two identical photodiodes situated as shown in Fig. 6-25*b* to *d*, then the amount of illumination received by each photodiode will depend on the level of the liquid. To be specific, let us assume that if the level is too high, as in Fig. 6-25*b*, both photodiodes receive an illumination of 1,300 fc. If the level is just right, as in Fig. 6-25*c*, photodiode P_1 receives an illumination of 2,000 fc and photodiode P_2 receives an illumination of 1,300 fc. Finally, if the level is too low, both photodiodes receive an illumination of 2,000 fc. Since the *v*-*i* curve of each photodiode changes with light illumination, it is clear that the voltage developed across the photodiode P_1 will not be the same in each case. Hence, if appropriate voltage-sensing devices such as relays are connected across the photodiode P_1 (for simplicity, let us assume the loading effect is negligible), then depending on the voltage across P_1, one relay will be actuated at a given time. This relay will in turn actuate an

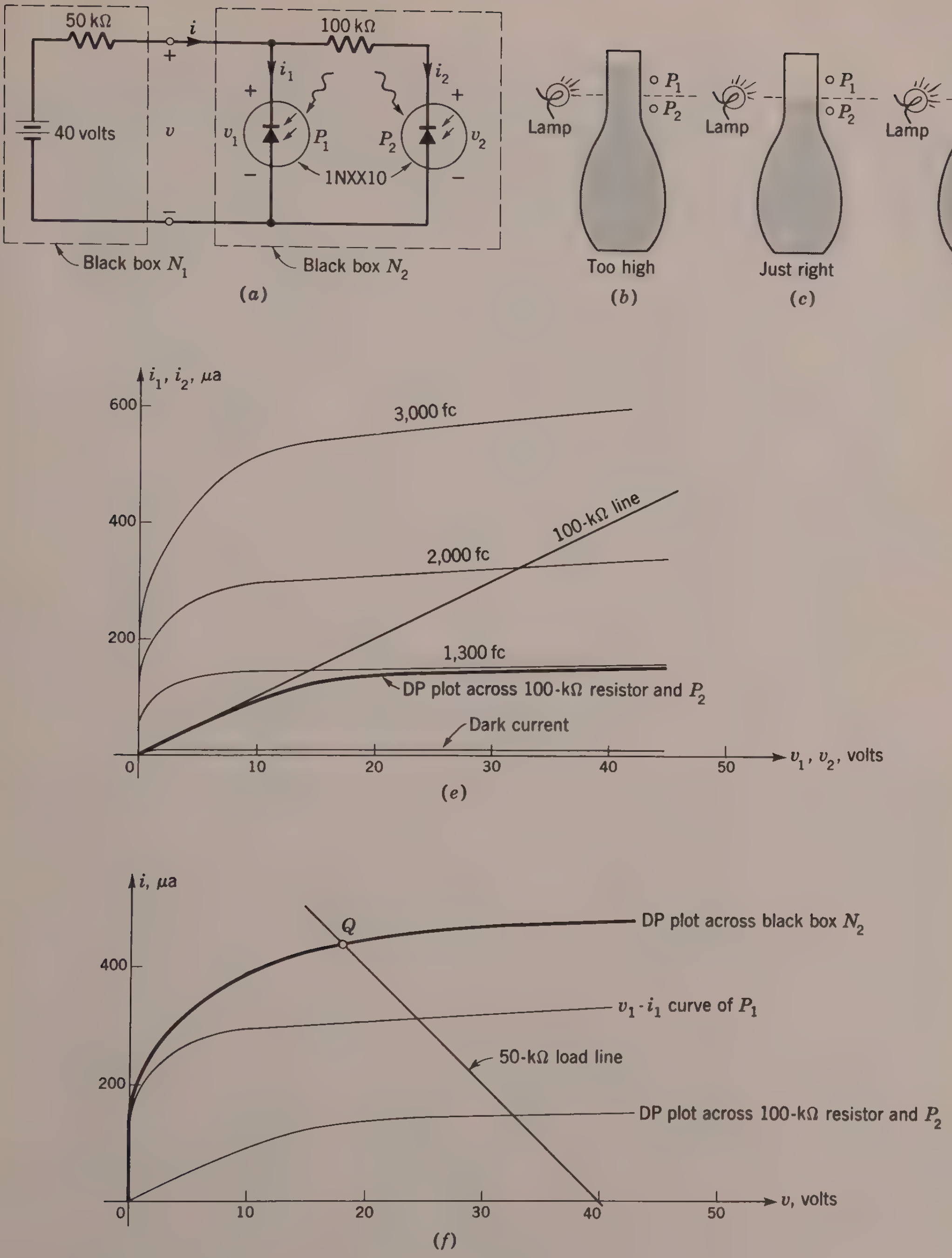


Fig. 6-25. A basic liquid-level sorting circuit and the graphical construction for determining the operating voltage corresponding to the case where the level is just right.

automatic mechanism to channel the bottle into the proper conveyor. In this way, only the bottle filled to the right level will arrive at the final sealing location; bottles not filled properly will be diverted into separate lines designed to correct the discrepancy.

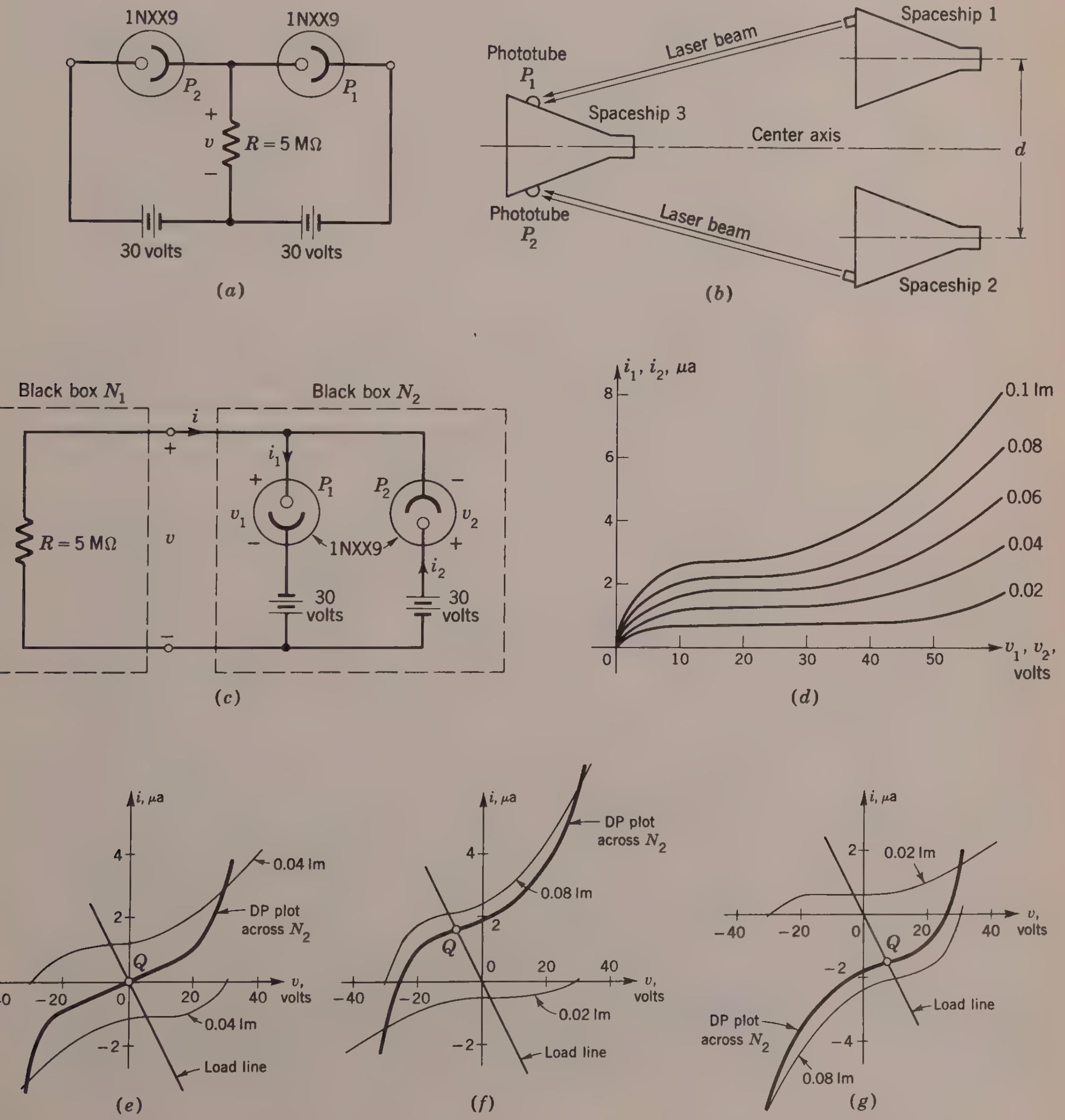
In order to choose the appropriate relays, we must solve for the operating point of the circuit shown in Fig. 6-25*a*. Since the voltage across P_1 is our only concern, it is obvious that we locate the two black boxes N_1 and N_2 as shown by the dotted line. Since the procedure is identical in each instance, let us consider only the one where the liquid level is just right. Referring to the characteristic curves of the photodiode type 1NXX10 in Appendix D, we locate the two curves corresponding to a light illumination of 1,300 and 2,000 fc. These two curves are drawn in light lines in Fig. 6-25*e*. The DP plot across the resistor-photodiode combination is obtained by adding the voltage coordinates of the 100-kΩ line and the v_2-i_2 curve of P_2 as shown in Fig. 6-25*e*. If we add the current coordinates of this curve with the v_1-i_1 curve of P_1 we obtain the DP plot across the black box N_2 as shown in Fig. 6-25*f*. The intersection Q between this DP plot and the load line is the operating point. Hence, the voltage corresponding to this case is $v = 18$ volts. Carrying out the same construction for the remaining cases, we find the voltage $v = 25.2$ volts when the level is too high and $v = 16.6$ volts when the level is too low.

The above principle has been utilized in the design of many instrumentation circuits. It is also used in the design of photometers for measuring light illumination; densitometers for measuring the density of liquid solutions; colorimeters for analyzing the color content of translucent materials and chemical solutions; and spectrophotometers for measuring the spectrum of a light source. It should be clear that we need not restrict ourselves to photodiodes. A phototube or a solar cell would do just as well. In fact, in the latter case, no external battery is necessary. Since the current and voltage furnished by a solar cell are very small, the price we pay for saving a battery is that a more sensitive monitoring device or circuit will be needed. This principle is not restricted to light-sensitive controlled resistors only. In many applications such as manometers, anemometers, flowmeters, and thermal conductivity detectors, the choice of temperature-sensitive controlled resistors such as thermistors is more suitable.

6-4-6 AUTOMATIC COMPARISON AND NULL DETECTION CIRCUIT

We shall now consider an extremely useful circuit for comparing the distance, position, direction, etc., between two objects as well as other physical quantities. A typical circuit of this nature is shown in Fig. 6-26*a*. This circuit has been proposed for achieving the three-spaceship flight formation scheme shown in Fig. 6-26*b*. This flight formation is part of an elaborate experiment proposed to qualify the spaceships for interplanetary expedition. The experiment calls for spaceship 1 to maintain a fixed distance *d* from

Fig. 6-26. A basic position-comparison and null-detection circuit and the graphical construction for determining the operating voltage corresponding to three different relative positions.



spaceship 2, and for spaceship 3 to move along an imaginary center axis between spaceships 1 and 2. The first requirement can be satisfied by simple radar control. Our problem, therefore, consists of maintaining the position of spaceship 3 along the center axis. To achieve this task, it is proposed that two laser beams of equal intensity be directed from spaceships 1 and 2 toward a phototube mounted on each side of spaceship 3, as shown in Fig. 6-26*b*. If spaceship 3 is exactly on the center line, both phototubes will receive equal light intensity. On the other hand, if spaceship 3 is closer to spaceship 1, phototube 1 will receive a higher light intensity than phototube 2 because the second laser beam has to travel a greater distance, thereby becoming more attenuated.[1] The opposite situation is encountered if spaceship 3 is closer to spaceship 2. Our problem is to find the voltage across the 5-MΩ resistor corresponding to different locations of spaceship 3 so that the astronaut inside would be able to steer his spaceship in accordance with the instrument reading.[2]

Since we are interested only in the voltage across the resistor, let us redraw the circuit into the black-box configuration shown in Fig. 6-26*c*. Our object is to find the DP plot across the black box N_2. For convenience, the characteristic curves for the phototube 1NXX9 are redrawn in Fig. 6-26*d*. Let us suppose that both phototubes receive a light illumination of 0.04 lm when the spaceship is on the center axis. The graphical construction corresponding to this case is shown in Fig. 6-26*e*. The resulting DP plot intersects the load line at the origin, and hence $v = 0$ for this case. The monitoring device must therefore indicate a *null reading* when spaceship 3 is on the center line.

Suppose, next, spaceship 3 is closer to spaceship 1 such that phototube 1 receives 0.08 lm while phototube 2 receives only 0.02 lm. The DP plot in this case is shown in Fig. 6-26*f*, and the voltage at the operating point is given by $v = -7.5$ volts. The opposite case where phototube 1 receives 0.02 lm and phototube 2 receives 0.08 lm is shown in Fig. 6-26*g*. By carrying out the same procedure for different light illuminations, a "flight-control curve" can be plotted indicating the relationship between the voltmeter reading and the relative location of spaceship 3. The astronaut can, therefore, steer his spaceship in accordance with this curve. Alternately, a control circuit can be designed to do this automatically.

It should be obvious by now that this principle can be utilized for designing many other control circuits, such as an airplane instrument-landing system. The phototubes in this case are mounted at the end of each wing, and the two light beams are directed from opposite sides of the runway. In all cases, a null reading indicates an optimum operation, and the control circuitry's job is to try to achieve a null by comparing the deviation from the desired position. Because of this interpretation, this class of circuit is called a *comparison and null-detection circuit.*

---

[1] It is assumed that the distance from one spaceship to another is so vast that the laser beam becomes sufficiently attenuated to register a significant unbalance in the light intensity received at the phototubes. The alert reader should recognize the contrived nature of this example. The important concept to learn here is the *balancing property* of the bridge phototube circuit. This property has been widely used in many more mundane applications.

[2] This resistor represents the input resistance of the monitoring device such as a voltmeter or other control circuitry.

A closer inspection of the circuit shown in Fig. 6-26*a* shows that it is a bridge circuit with the "sensing" element located at opposite arms. The ability to detect any unbalance between the two arms is a basic property that makes the bridge circuit so useful for comparison purposes. If we replace the phototubes by two thermistors, the result is, of course, a sensitive temperature-comparison circuit. There is an almost endless variety of useful circuits that are based primarily on this basic bridge configuration. However, since the principle is the same, we shall not labor any longer to discuss them. After all, one of the merits of the black-box approach is the possibility of extracting the general concepts from many seemingly unrelated circuits.

Exercise 1: (*a*) Using the v-i curves for the junction diode type 1NXX2 and the zener diode type 1NXX3, obtain the exact DP plot of the concave resistor shown in Fig. 6-20*a* with R equal to 50 Ω, 100 Ω, 500 Ω, and 1 kΩ. (*b*) Compare each DP plot with that obtained from piecewise-linearizing the v-i curves. (*c*) Formulate a simple rule to correct the error resulting from the piecewise-linearization.

Exercise 2: Show how a convex resistor may be synthesized by a concave resistor and a 45° reflector or gyrator.

Exercise 3: (*a*) Verify the DP plot shown in Fig. 6-24*c*. (*b*) Find the new DP plot if the valley current of each tunnel diode is less than the peak current but greater than the valley current of the preceding tunnel diode.

Exercise 4: (*a*) Find the operating voltage of the liquid-level sorting circuit in Fig. 6-25*a* when the level is too high and when it is too low. (*b*) Why is the 100-kΩ resistor needed? How should its value be chosen?

Exercise 5: Using the solar cell type 1NXX11, design a liquid-level sorting circuit without batteries for carrying out the same task as in Fig. 6-25.

Exercise 6: To improve reliability, it is desirable that the operating voltage in each case of the sorting circuit in Fig. 6-25 differ by a relatively large amount. Show that this can be achieved by replacing the 50-kΩ linear resistor by a nonlinear resistor with an appropriate v-i curve. Specify a v-i curve so that the operating voltage in Fig. 6-25*c* differs from the operating voltage in Fig. 6-25*b* and *d* by an equal amount.

Exercise 7: Verify the graphical construction shown in Fig. 6-26*e* to *g*.

Exercise 8: In practice, the two phototubes shown in Fig. 6-26*a* may not be identical. To achieve a null reading when spaceship 3 is on the center axis, a resistance may be inserted in series with each phototube. Show how the values must be chosen to achieve a null. Assume that the v-i curves of phototube 2 are displaced in the vertical direction from the original v-i curve by 1 μa.

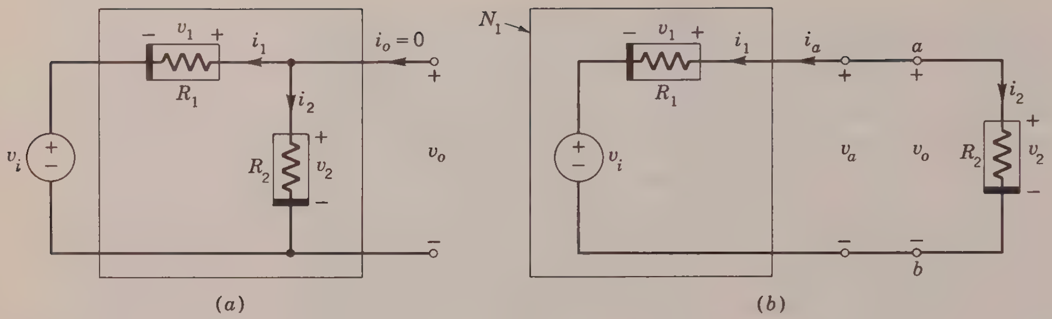


Fig. 6-27. To determine the TC plot of the nonlinear voltage divider, we redraw the network into the form of basic configuration 1.

6-5 GRAPHICAL DETERMINATION OF TC PLOTS OF SERIES-PARALLEL NETWORKS

Here we shall consider how to determine the TC plots of series-parallel networks containing only two-terminal resistors and dc sources. Although there are four distinct types of TC plots, the graphical procedures that we are about to study are applicable to all. The method consists of treating each series-parallel network as a cascade connection of two basic network configurations, the nonlinear voltage divider and the nonlinear current divider.

6-5-1 NONLINEAR VOLTAGE DIVIDER

The circuit shown in Fig. 6-27*a* is called a *nonlinear voltage divider* because the input voltage divides, in accordance with some nonlinear relationship, between the two nonlinear resistors. To determine the v_o-vs.-v_i TC plot, we observe that $i_o = 0$, and hence the network can be redrawn into the form of basic configuration 1, as shown in Fig. 6-27*b*.[1] Since the v_o-vs.-v_i TC plot is simply the locus of the operating-point voltage $v_o = v_2$ corresponding to each value of the input voltage v_i, it can be easily found by the load-line method with R_2 as the load, provided the DP plot across the black box N_1 can be found for each value of v_i. Since the effect of the voltage source v_i is to translate the v_1-i_1 curve of R_1 along the voltage axis, it is convenient to cut a template of the v_1-i_1 curve and translate it by a distance equal to each value of v_i along the v_a axis, thereby generating a family of parallel-displaced v_a-i_a curves. The intersections between this family of curves and the load line associated with R_2 constitute a set of points in the v_o-vs.-v_i plane. The curve passing through these points is the desired TC plot. For future reference we shall refer to the above procedure as the *template method.*

---

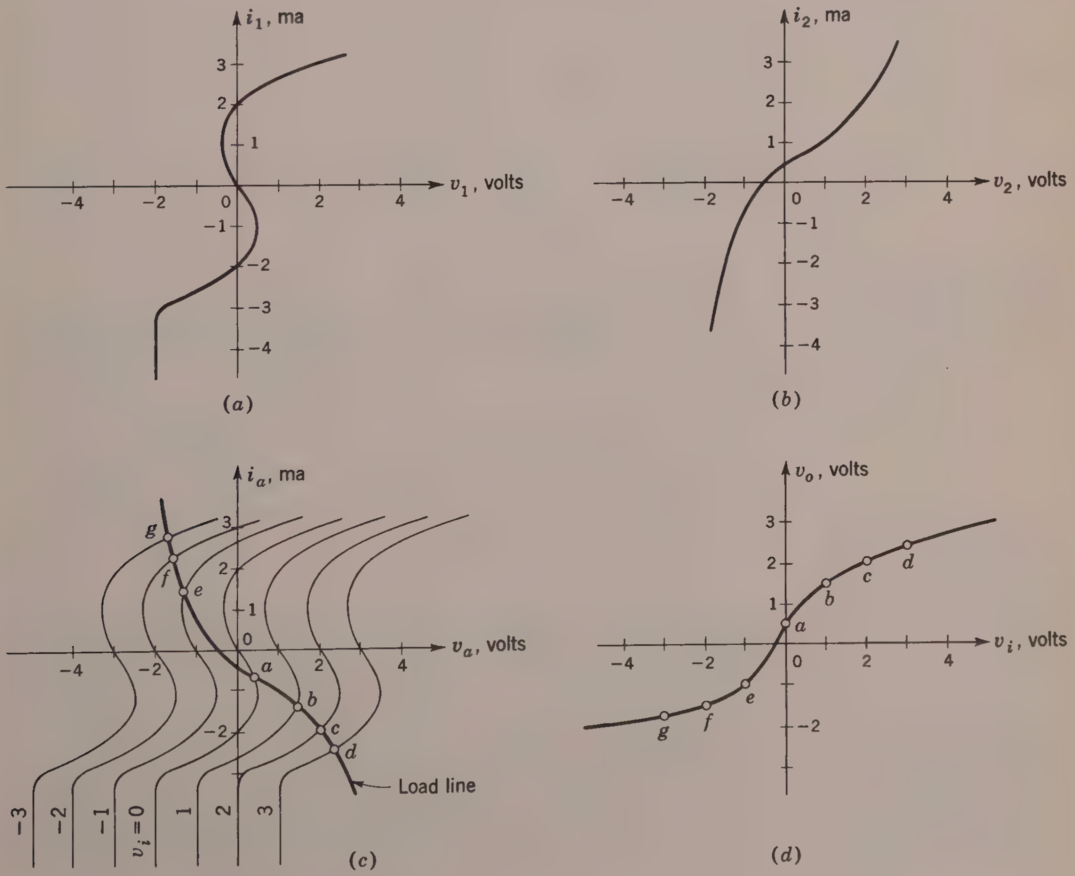
[1] Recall that the definition of the TC plot implies the condition $i_o = 0$, that is, no loading effects.

EXAMPLES

1. Suppose the resistors R_1 and R_2 are characterized by the curves shown in Fig. 6-28*a* and *b*, respectively. The family of v_a-i_a curves obtained by the template method is shown in Fig. 6-28*c*. The TC plot is finally obtained by plotting the coordinates v_i and $v_a = v_o$ of the points of intersection in the v_i-v_o plane as shown in Fig. 6-28*d*.

We shall now present another graphical method which is more amenable to computer implementation. This is the *graphical elimination method.* To apply this method, it is more convenient to redefine the reference voltage polarity and the reference current direction of resistor R_1 in Fig. 6-27*a* as shown in Fig. 6-29*a*. Since $i_o = 0$, we can redraw the voltage divider into

Fig. 6-28. The graphical procedure for determining the TC plot of a resistive nonlinear voltage divider by the template method.



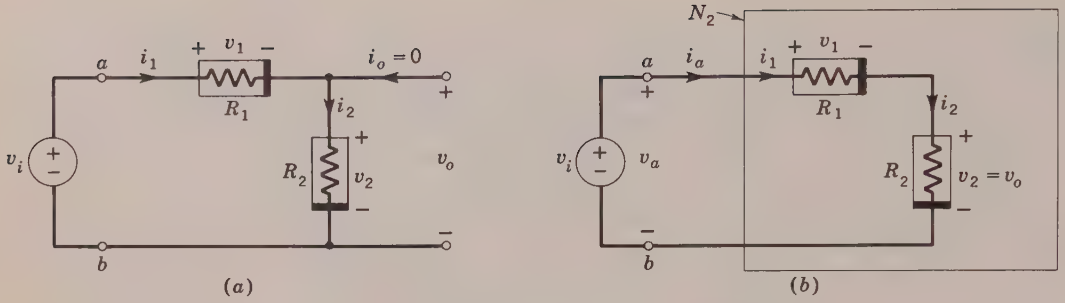


Fig. 6-29. To apply the graphical elimination method, the two nonlinear resistors must be redrawn into the back-to-front series configuration.

the form shown in Fig. 6-29*b*. The DP plot of the black box N_2 can be easily obtained by the series-combination technique and represents a curve in the v_a-i_a plane. Since $v_a = v_i$ and $i_a = i_2$, we can relabel the v_a axis of the DP plot by v_i, and the i_a axis by i_2. In other words, the same curve Γ representing the DP plot of N_2 also represents the relationship between i_2 and v_i. Similarly, since $v_2 = v_o$, the v_2-i_2 curve of the nonlinear resistor R_2 can also be considered as a relationship between i_2 and v_o. Since both the v_i-vs.-i_2 curve and the v_o-vs.-i_2 curve involve the same variable i_2, this variable can be eliminated graphically to obtain a relationship between v_i and v_o, which is the desired TC plot.[1] As a summary, the graphical elimination method consists of two steps. The first step consists of determining the DP plot of the two resistors in series (with due attention given to the reference direction and polarity of R_1). The second step consists of performing a graphical elimination procedure between this DP plot and the v_2-i_2 curve of R_2.†

2. For comparison purposes, let us apply the graphical elimination method to the same circuit considered earlier in Example 1. First, we must redefine the reference polarity and direction for R_1. The effect of doing this is that the v_1-i_1 curve given in Fig. 6-28*a* must be rotated by 180° with respect to the origin, as shown in Fig. 6-30*a*. The v_2-i_2 curve remains the same and is drawn beside the v_1-i_1 curve as shown in Fig. 6-30*b*. The v_a-i_a DP plot is obtained as shown in Fig. 6-30*c*. To eliminate the variable i_2, we first draw a 45° reflection line as shown in Fig. 6-30*d*. Next, we draw the set of axes v_i and v_o in line with the horizontal axis of Fig. 6-30*d* and the vertical axis of Fig. 6-30*c*, as shown in Fig. 6-30*e*. The procedure now consists of drawing a vertical guideline through any value $v_i = k$ until it intersects the v_a-i_a DP plot at some point α. A horizontal guideline is next drawn through α until it intersects the v_2-i_2 curve

---

[1] The graphical elimination procedure for eliminating the common variables between two curves is given in Appendix B.

† A simple example will show that if the resistor R_2 is voltage-controlled, or multivalued, one must be careful to pick only "corresponding points" in the elimination procedure. Otherwise, extraneous segments may be introduced.

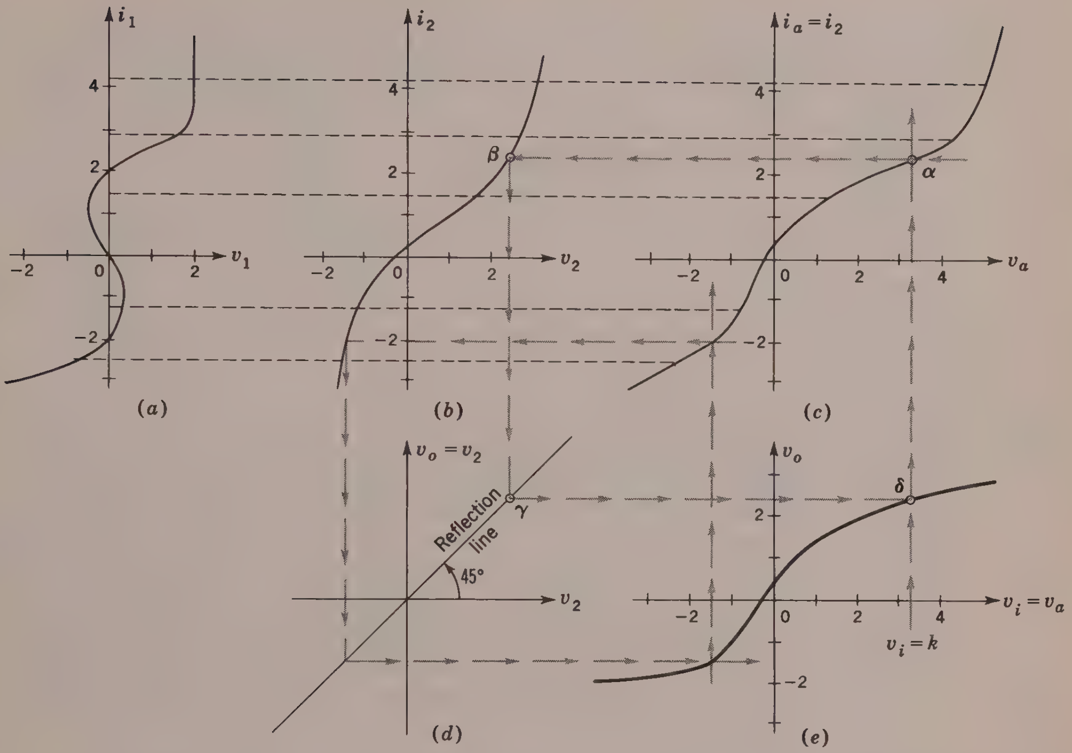


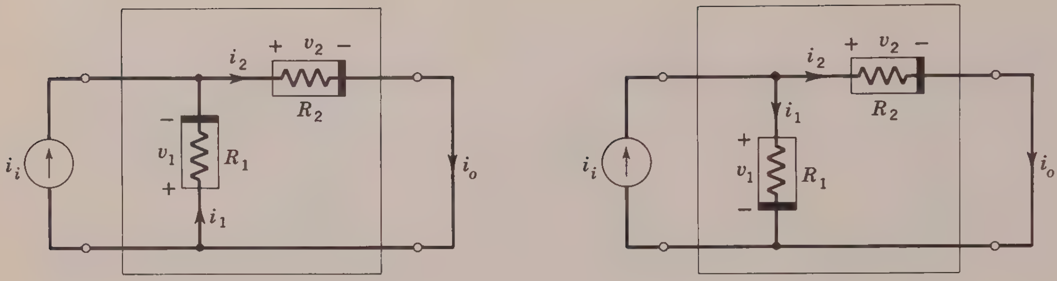
Fig. 6-30. The graphical procedure for determining the TC plot of a resistive nonlinear voltage divider by the graphical elimination method.

at some point β. From β we drop a vertical guideline until it intersects the reflection line at γ. The intersection δ of the horizontal guideline through γ with the initial vertical guideline locates one point on the TC plot. If this procedure is repeated a sufficient number of times, we would obtain the TC plot shown in Fig. 6-30*e*. Observe that this TC plot is identical with that obtained earlier in Fig. 6-28*d* by the template method, as it should be.

A comparison between the two methods shows that if both the v_1-i_1 and the v_2-i_2 curves are monotonic, the graphical elimination method is easier. However, if the v_1-i_1 and the v_2-i_2 curves are not monotonic, it may take more time to find the DP plot than to cut a template and draw the family of curves. Moreover, the template method gives better insight into the effect of the v_2-i_2 curve of R_2 on the resulting TC plot. In any case, however, the graphical elimination method is much easier to carry out on a computer.[1]

---

[1] A computer program for implementing the graphical elimination method can be easily written. All it takes is a graphical elimination subroutine.



(a) (b)

Fig. 6-31. The basic form of the nonlinear current-divider network appropriate for the template method (left) and the graphical elimination method (right).

6-5-2 NONLINEAR CURRENT DIVIDER

The network shown in Fig. 6-31*a* and *b* is called a *nonlinear current divider* because the input current i_i divides, in accordance with some nonlinear relationship, between the two nonlinear resistors. The reader will recognize the network on the left as the dual of the nonlinear voltage divider shown in Fig. 6-27*a*, and hence the i_o-vs.-i_i TC plot can be obtained by the dual of the template method. Similarly, the network on the right is the dual of the nonlinear voltage divider shown in Fig. 6-29*a*, and hence the i_o-vs.-i_i TC plot can be obtained by the dual of the graphical elimination method.

6-5-3 SERIES-PARALLEL NONLINEAR NETWORKS

The v_o-vs.-v_i or the i_o-vs.-i_i TC plot of a series-parallel nonlinear network can be determined by a successive application of the methods described for the nonlinear voltage divider or the nonlinear current divider, respectively. We shall illustrate the procedure for the v_o-vs.-v_i case by an example, leaving the dual case for the problems.

EXAMPLE

Suppose it is required to determine the v_o-vs.-v_i plot of the series-parallel nonlinear network shown in Fig. 6-32*a*. The first thing to do is to reduce this network into the simpler *ladder* network shown in Fig. 6-32*b* by replacing resistors R_1, R_2, and R_3 by R_{14}; resistors R_4 and R_5 by R_{15}; resistors R_6 and R_7 by R_{16}; resistors R_8 and R_9 by R_{17}; and resistors R_{10}, R_{11}, and R_{12} by R_{18}. Each of these new resistors is, of course, characterized by the DP plot of the corresponding group of resistors.

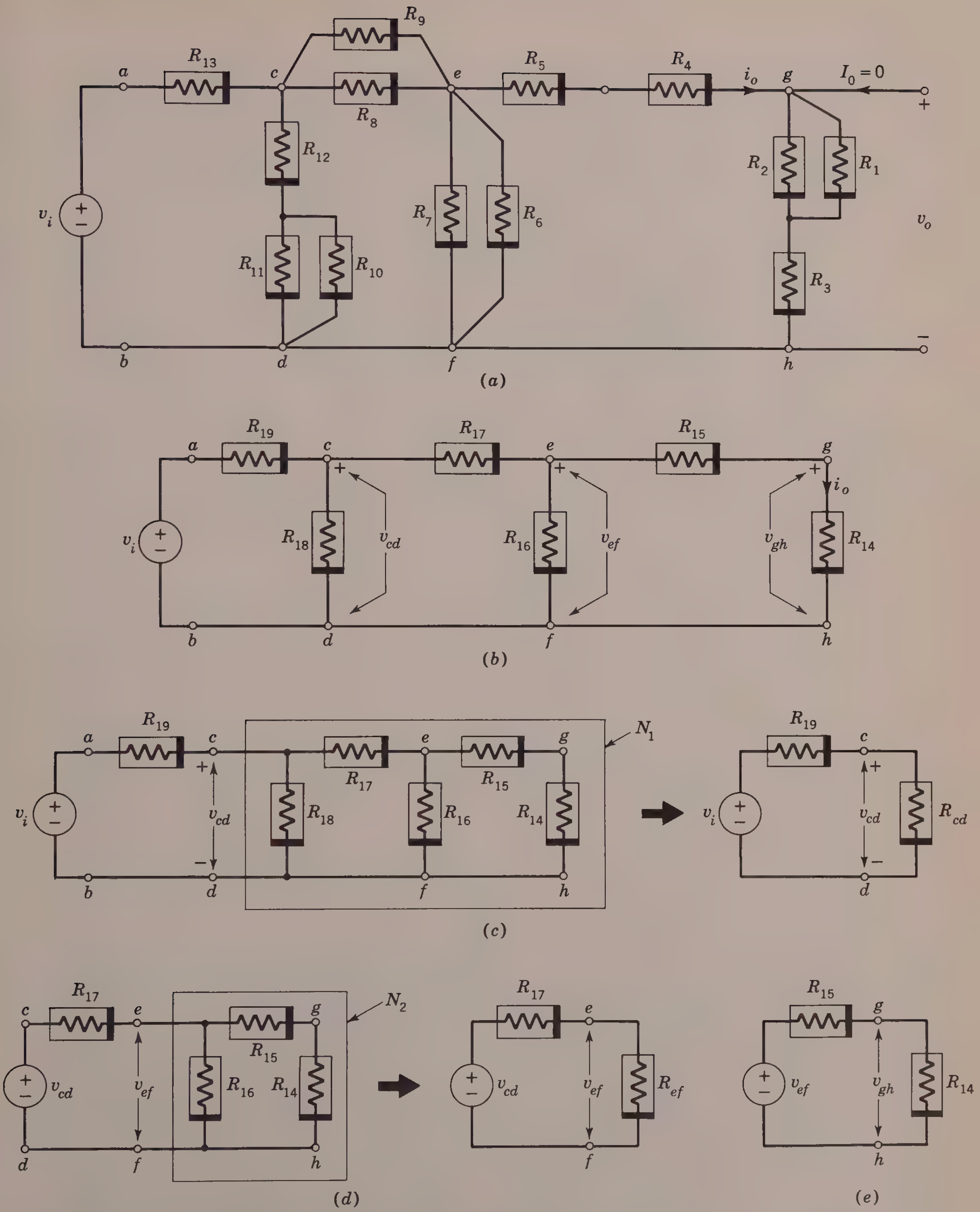


Fig. 6-32. The various reduction procedures for determining the v_o-vs.-v_i plot of a series-parallel nonlinear network.

The next step is to determine a number of related TC plots. First, we create a soldering-iron entry across terminals *c-d* in Fig. 6-32*b* and replace the black box N_1 by resistor R_{cd} as shown in Fig. 6-32*c*. The result is a nonlinear voltage divider, and hence the v_{cd}-vs.-v_i TC plot can be easily obtained.

The next step is the crucial one. It is based on the observation that if the relationship between v_{ef} across terminals *e-f* in Fig. 6-32*b* and v_{cd} can be found, then the curve v_{ef}-vs.-v_{cd} can be "composed" with the v_{cd}-vs.-v_i plot to obtain the v_{ef}-vs.-v_i TC plot.[1] The relationship between v_{ef} and v_{cd} can be found by first replacing that part of the network to the left of terminals *c-d* in Fig. 6-32*b* by a voltage source with a terminal voltage v_{cd}, as shown in Fig. 6-32*d*. We can do this because v_{cd} is now the *independent* variable for the v_{ef}-vs.-v_{cd} curve. We can now repeat our earlier technique by creating a soldering-iron entry across terminals *e-f* and replace the black box N_2 by resistor R_{ef} as shown in Fig. 6-32*d*. The result is again a voltage divider, and the v_{ef}-vs.-v_{cd} plot can be found.

To be specific, let us suppose the v_{cd}-vs.-v_i plot obtained earlier and the v_{ef}-vs.-v_{cd} plot just obtained are given, respectively, by the curves shown in Fig. 6-33*a* and *b*. If we denote these curves by the functional notations $v_{cd} = g(v_i)$ and $v_{ef} = f(v_{cd})$, respectively, then the composition procedure is simply a systematic method for finding the function of another function, namely, $v_{ef} = f(g(v_i))$. One simple graphical procedure for doing this is shown in Fig. 6-33*c*. The v_{ef}-vs.-v_{cd} plot and the v_{cd}-vs.-v_i plot are drawn in the upper left-hand and the lower right-hand corners, respectively. A 45° reflection line is drawn with its vertical axis in line with the v_{ef} axis and its horizontal axis in line with the v_i axis. Hence, starting at any value $v_i = k$, a vertical guideline is dropped until it intersects the v_{cd}-vs.-v_i curve at some point α. A horizontal guideline through α is drawn to the left until it intersects the reflection line at β. From β a vertical guideline is drawn until it intersects the v_{ef}-vs.-v_{cd} plot at γ. Finally, the point δ obtained from the intersection between the horizontal line through γ and the initial vertical line through α is a point on the v_{ef}-vs.-v_i curve. The other points on this curve are obtained in a similar manner.[2]

The v_{ef}-vs.-v_i plot gives us the voltage v_{ef} corresponding to each value of the input voltage v_i. If we can find the voltage v_{gh} corresponding to each value of v_{ef}, then, clearly, we can find the output voltage $v_o = v_{gh}$ corresponding to each input voltage v_i. Let us replace the network to the left of nodes *e-f* by a voltage source v_{ef} as shown in Fig. 6-32*e*. The result is again a voltage

---

[1] The reader should recall the basic composition technique described in Sec. 2-4. The graphical composition procedure is discussed in Appendix B.

[2] For complex, multivalued curves, this graphical composition procedure can be easily implemented by a computer subroutine.

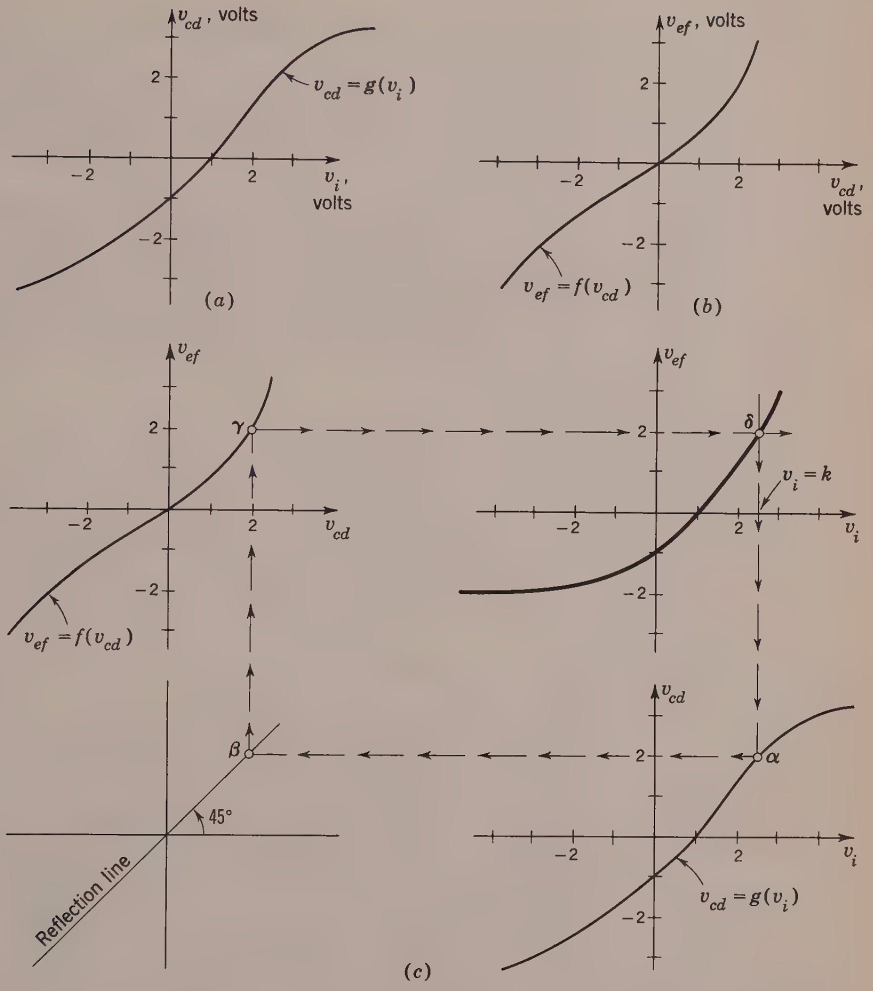


Fig. 6-33. The graphical composition procedure for obtaining the function of another function.

divider, and the v_{gh}-vs.-v_{ef} plot can be readily found. By similar reasoning, the graphical composition procedure can now be applied to the v_{gh}-vs.-v_{ef} plot and the v_{ef}-vs.-v_i plot to obtain the v_{gh}-vs.-v_i plot, which is the desired TC plot since $v_{gh} = v_o$. It should now be clear that no matter how complicated the network, the v_o-vs.-v_i plot of any series-parallel nonlinear resistive network can always be found by repeated application of these techniques.

Exercise 1: Let the resistor R_1 in Fig. 6-27*a* be characterized by the equation $i_1 = v_1^3 + 3v_1^2 - 2$ and the resistor R_2 be characterized by $i_2 = v_2^3$. (*a*) Find the v_o-vs.-v_i TC plot by the template method. (*b*) Repeat (*a*) by the graphical elimination method.

Exercise 2: Let the resistor R_1 in Fig. 6-31*a* be characterized by the equation $v_1 = i_1^3 + 3i_1^2 - 2$ and the resistor R_2 be characterized by $v_2 = i_2^3$. (*a*) Find the i_o-vs.-i_i TC plot by the template method. (*b*) Repeat (*a*) by the graphical elimination method.

Exercise 3: (*a*) Find the TC plot between the varistor voltage and the input voltage of the voltage-stabilizer network shown in Fig. 4-7*a*. (*b*) Find the varistor voltage when the input voltage is 200 volts. Verify that this answer checks with that obtained earlier in Sec. 4-5 by the iteration method. (*c*) Simplify this circuit by replacing the voltage source and the two linear resistors by an equivalent circuit containing one voltage source and one resistance, and then find the TC plot by the template method.

6-6 SOME PRACTICAL APPLICATIONS OF TC PLOTS

There are many electronic circuits designed specifically to transform an input signal into some desirable output signal. These are called *signal-processing circuits.* We shall now apply the TC plot concept to analyze this class of circuits.

6-6-1 HALF-WAVE RECTIFIER

The term rectifier was introduced earlier in Sec. 1-6-5 as pertaining to the idealized circuit shown in Fig. 1-12. We were forced to choose an ideal diode at that time so that we could obtain a qualitative behavior of the circuit function. We are now ready to perform an exact graphical analysis of the same circuit with realistic devices. This time, let us choose the vacuum diode type 1NXX1 in the circuit shown in Fig. 6-34*a*. The *v*-*i* curve of this diode is given in Appendix D and is reproduced in Fig. 6-34*b* for convenience. The given reference polarity and direction are already in the right form for the graphical elimination method. However, since we are interested not only in the TC plot but also in how the value of the load resistor *R* affects the TC plot, we shall choose the template method. This means that we must remember to rotate the *v*-*i* curve by 180° with respect to the origin before drawing the family of v_a-i_a curves as shown in Fig. 6-34*c*. The TC plot corresponding to the 2-kΩ load line is shown in the upper left-hand corner of Fig. 6-34*d*. If we apply a sinusoidal voltage as shown in the lower right-hand corner of this figure, then a graphical composition between the TC plot $v_o = f(v_i)$ and the input voltage $v_i = v_i(t)$ would give the output waveform $v_o = f(v_i(t))$, as shown in the upper right-hand corner of Fig. 6-34*d*. A com-

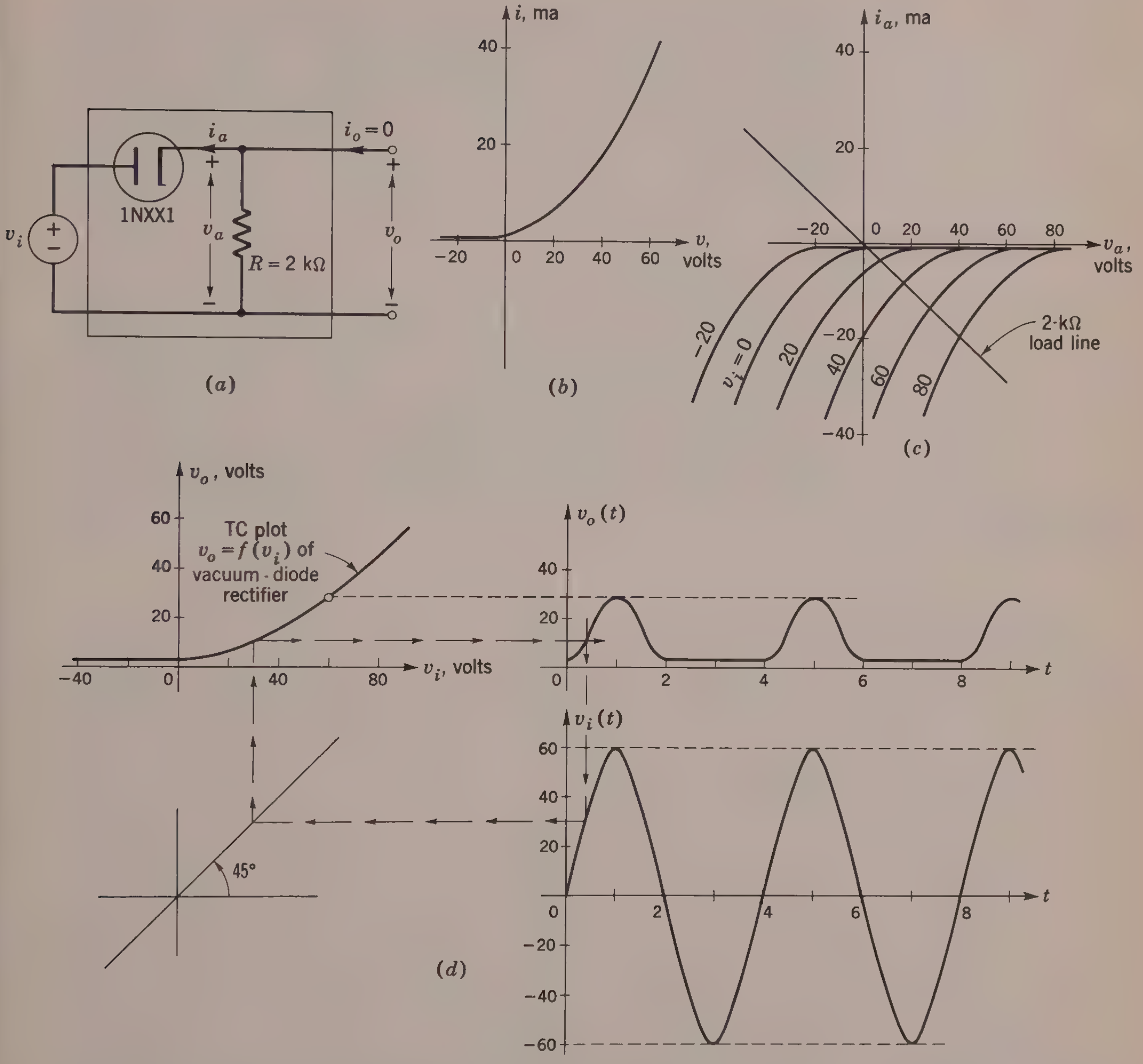


Fig. 6-34. A complete graphical analysis of the half-wave rectifier.

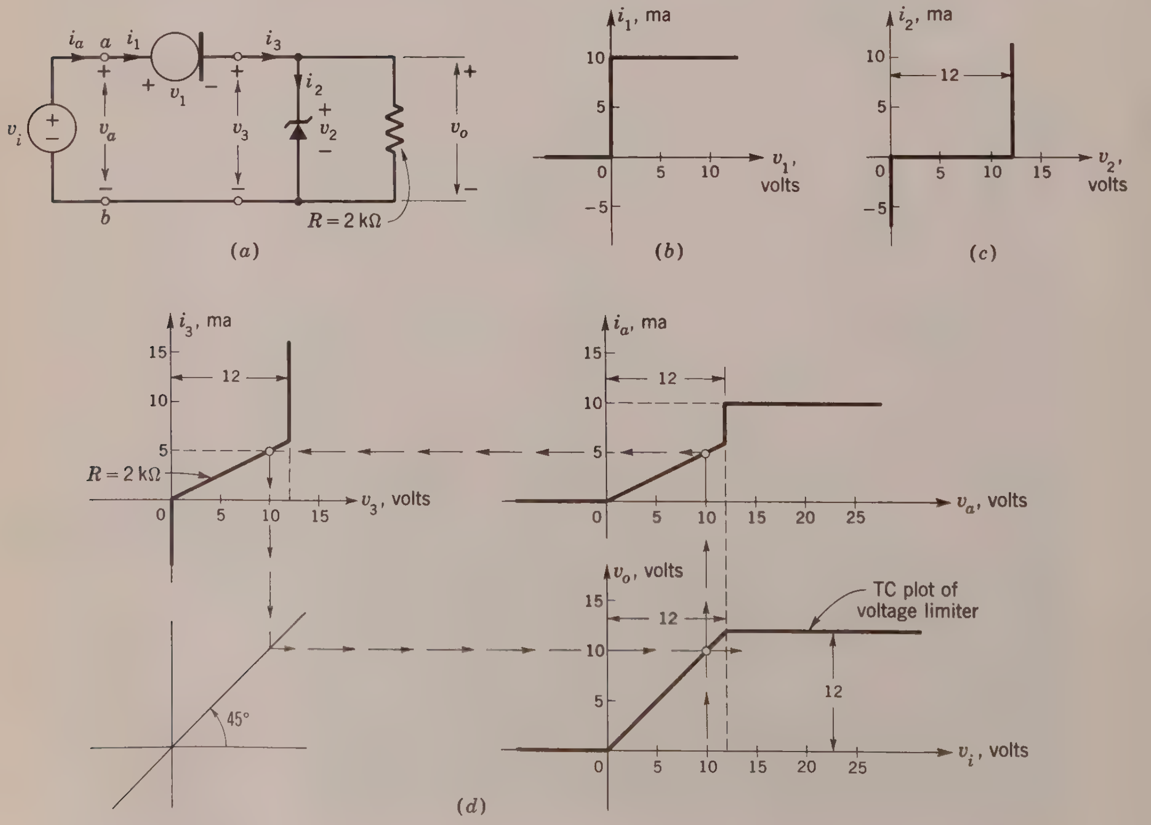
parison between this exact solution and the idealized solution in Fig. 1-12 shows that the rectification is not perfect because the output voltage is not exactly zero during alternate half-cycles. Moreover, the peak output voltage is smaller than the peak input voltage.

In practice, the load resistor R varies, and hence it is desirable to know what effect it has on the TC plot. One advantage of the template method is that only the load line varies, and hence it is relatively simple to obtain the desired information. One could, for example, obtain several TC plots with the value of R as a parameter. For this simple example, this information can be obtained by inspection of the curves in Fig. 6-34c and the movement of the points of intersection as the load line varies. For example, it is easily seen that the peak output voltage decreases as the value of R decreases.

6-6-2 VOLTAGE LIMITER

Consider the circuit shown in Fig. 6-35a. The v_1-i_1 curve of the constant-current diode is shown in Fig. 6-35b, and the v_2-i_2 curve of the zener diode is shown in Fig. 6-35c. A load resistor R is normally connected across the output terminal, thereby violating the requirement that the load current be zero in a voltage divider. However, this loading effect can always be taken into consideration by combining the load resistor with the zener diode as one

Fig. 6-35. Graphical analysis of a voltage limiter.



black box. The resulting v_3-i_3 curve of this parallel combination is shown in the upper left-hand corner of Fig. 6-35*d*. To the right of this is the DP plot across terminals *a-b*. The TC plot obtained by the graphical elimination method is shown in the lower right-hand corner of Fig. 6-35*d*.

An examination of this TC plot shows that $v_o = v_i$ for all $0 \leq v_i \leq 12$, and $v_o = 12$ for all $v_i \geq 12$. Since the circuit limits the output voltage from exceeding 12 volts, it is called a *voltage limiter.* A more careful examination of the graphical construction reveals that the same TC plot would be obtained for all $R \geq 12/10 = 1.2$ kΩ. This circuit has many useful applications. It is widely used for overload protection because no matter how the input voltage varies, the load voltage will never exceed the limiting voltage. It is also used as a voltage regulator for furnishing a constant output voltage $v_o = 12$ volts even if the input voltage v_i varies (assuming $v_i \geq 12$ volts). While there are many other circuits that can be designed to function as voltage limiters, this circuit has the advantage that no batteries are needed.

6-6-3 PULSE COMPRESSOR

The circuit shown in Fig. 6-36*a* is called a pulse compressor. To understand what it does, let us first determine its TC plot. For simplicity, we have represented the tunnel diode v_1-i_1 curve by three piecewise-linear segments as shown in Fig. 6-36*b*. In the usual case, the values of R_1 and R_2 are chosen so that $R_1 + R_2 = R$, where R is the magnitude of the negative resistance of the tunnel diode. Using the graphical elimination method, we obtain the TC plot of this circuit as shown in Fig. 6-36*c*. The operation performed by this TC plot is readily seen by considering the input signal waveform $v_i(t)$ shown in the lower right-hand corner of Fig. 6-36*d*. The corresponding output waveform $v_o(t)$ is readily obtained by the graphical composition procedure and is shown in the upper right-hand corner of Fig. 6-36*d*. Notice that the upper portion of the output signal remains the same as that of the input signal; only its height is reduced, hence the name *pulse compressor.* This type of circuit is useful in processing video signals of color television and radar sets, where only the waveform on top of each pulse is important. One could imagine this information-bearing signal to be riding on top of a rectangular pulse as indicated by the dotted lines in the lower right-hand corner of Fig. 6-36*d*. If this signal is fed directly into an amplifier, the larger pulses would generally drive the amplifier into the saturation region,[1] and hence the sig-

---

[1] Recall the characteristics of the operational amplifier in Sec. 3-6-2.

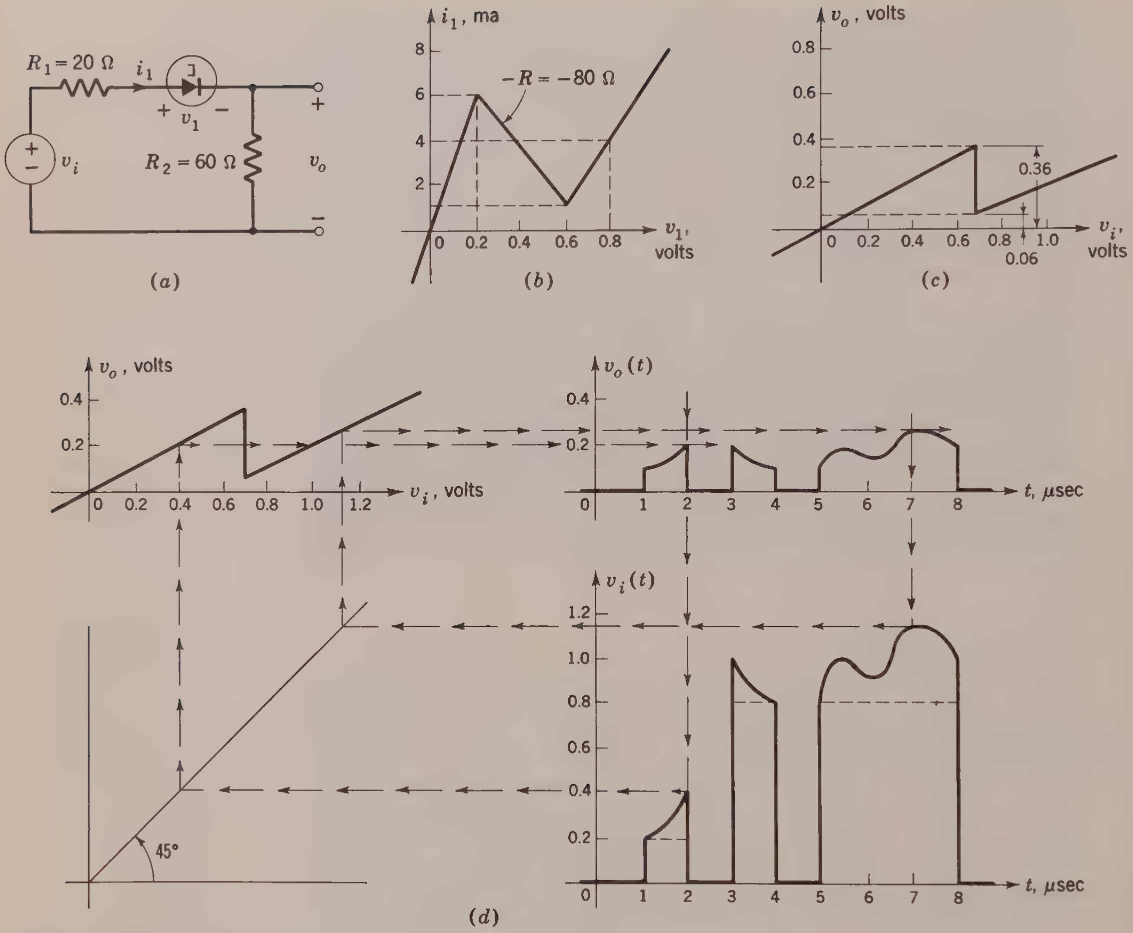


Fig. 6-36. Graphical analysis of a pulse compressor and its effect on the height of a pulse signal.

nal will not be amplified properly. Since we are interested only in the signal on top of the pulse, the output signal from the pulse compressor can be applied to the amplifier, thereby obtaining an output signal without distortion. This statement is, of course, valid only if the projection of that portion of the signal of interest in the graphical construction of Fig. 6-36*d* lies either to the left or to the right of the vertical segment of the TC plot.

6-6-4 VERTICAL-PULSE GENERATOR

In many applications, especially in computers, it is desirable to generate pulses which rise and fall in zero time. Unfortunately, this goal can be realized only approximately in practice.[1] The square-wave generator described in Sec. 3-6-1 represents one such realization. The quality of the generated pulses in this case depends on the frequency and amplitude of the input signal. The circuit shown in Fig. 6-37*a* has the advantage that it is independent of the frequency.

---

[1] As will be seen in Part 3 of this book, the presence of parasitic elements prevents the generation of such pulses.

Corresponding to the tunnel diode v_2-i_2 curve shown in Fig. 6-37*b*, we obtain by the template method (shown in Fig. 6-37*c* corresponding to a particular value of R_1) the TC plot shown in the upper left-hand corner of Fig. 6-37*d*. The output signal $v_o(t)$ corresponding to a sinusoidal input signal is shown in the upper right-hand corner of Fig. 6-37*d*. Observe that the output waveform contains *almost* vertical pulses, independent of the frequency of the input signal. This part of the waveform is the important one. In most applications, the waveform between the vertical edges of each pulse is not important and is usually "flattened" by a circuit called a *clipper,* to be discussed in Chap. 9.

So far, we have restricted our analysis to TC plots. In many electronic circuits, it is just as important to obtain the *power-transfer plot* with respect to the low-power nonlinear devices so that the maximum allowable input voltage may be found. We have already demonstrated in Chap. 5 how to obtain this relationship. Let us apply the procedure described in Sec. 5-4 to determine the power-transfer plot of the tunnel diode. First, we notice that the power across the tunnel diode is given by

$$\begin{aligned} p &= i_2(v_2)v_2 \\ &= i_2(v_o)v_o \end{aligned} \tag{6-7}$$

If we denote the TC plot by the functional notation $v_o = f(v_i)$, then Eq. (6-7) becomes

$$p = i_2(f(v_i))f(v_i) \tag{6-8}$$

The power-transfer plot can, therefore, be found by performing first a graphical composition between the v_2-i_2 curve of the tunnel diode and the $v_o = f(v_i)$ TC plot.[1] The resulting $i_2(f(v_i))$ curve is shown in the upper right-hand corner of Fig. 6-37*e*. The power-transfer plot can now be obtained by multiplying the ordinates of this curve and that of the TC plot, corresponding to each value of v_i. The resulting power-transfer plot is shown in the lower right-hand corner of Fig. 6-37*e*. From this curve, we can immediately find the maximum input voltage $E_{\max}$ that can be applied without exceeding the given maximum power rating $P_{\max}$ of the tunnel diode.

---

[1] A simpler equivalent method for obtaining this relationship is given in Exercise 5.

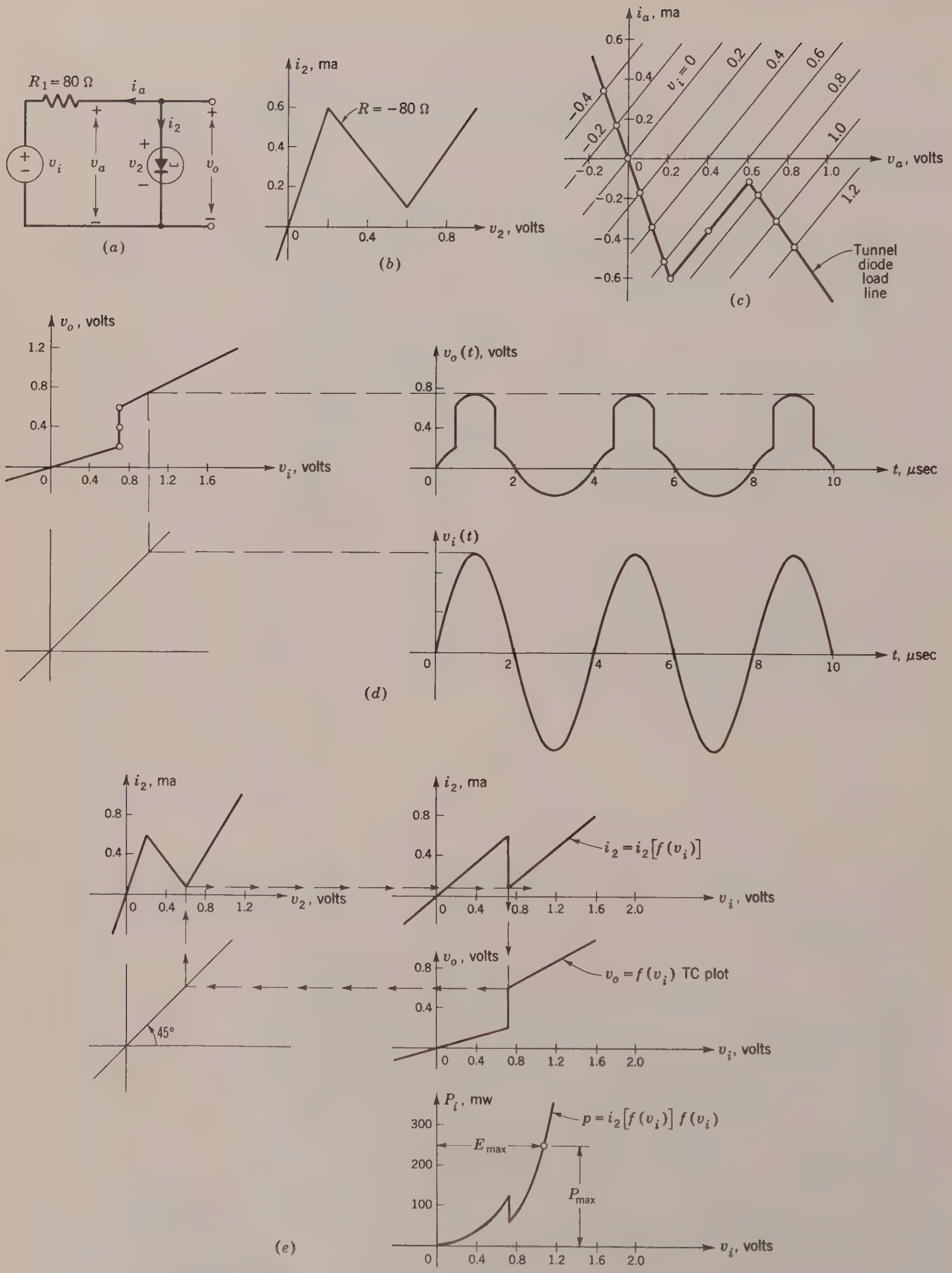


Fig. 6-37. A complete graphical analysis of the vertical pulse generator.

Exercise 1: (*a*) Find the TC plot of the rectifier circuit shown in Fig. 6-34*a* by the graphical elimination method. (*b*) Plot the relationship between the peak output voltage as a function of the load resistance R when the peak input voltage is 40 volts.

Exercise 2: (*a*) Find the TC plot of the voltage limiter by the template method. (*b*) Show why it is that the same TC plot in Fig. 6-35*d* would result for all values of $R \geq 1.2$ kΩ. (*c*) Replace the constant-current diode in Fig. 6-35*a* by a 1-kΩ linear resistor and find the TC plot. Compare this TC plot with that of the original circuit and find the advantage in using a constant-current diode.

Exercise 3: (*a*) A symmetrical voltage limiter can be designed by adding one constant-current diode in parallel (back to front) with the original constant-current diode, and a zener diode in series (back to back) with the original zener diode. Find the TC plot of this modified circuit and interpret the functions performed by this circuit. (*b*) The constant-current diode in this modified circuit may be replaced by an FET and a junction diode. Show how this can be achieved.

Exercise 4: (*a*) Obtain the TC plot of the pulse-compressor circuit shown in Fig. 6-36*a* by the template method and by the graphical elimination method. (*b*) Which method is easier for this circuit? (*c*) Based on your observation from (*b*), can you state some ground rule as to which method is preferred in a particular circuit?

Exercise 5: (*a*) Find the TC plot of the vertical-pulse generator circuit in Fig. 6-37*a* by the graphical elimination method. (*b*) Verify that the $i_2(f(v_i))$ curve in Fig. 6-37*e* is identical with the DP plot across resistor R_1 and the tunnel diode in series. (*c*) Show that the simpler procedure described above is valid for any nonlinear voltage-divider circuit.

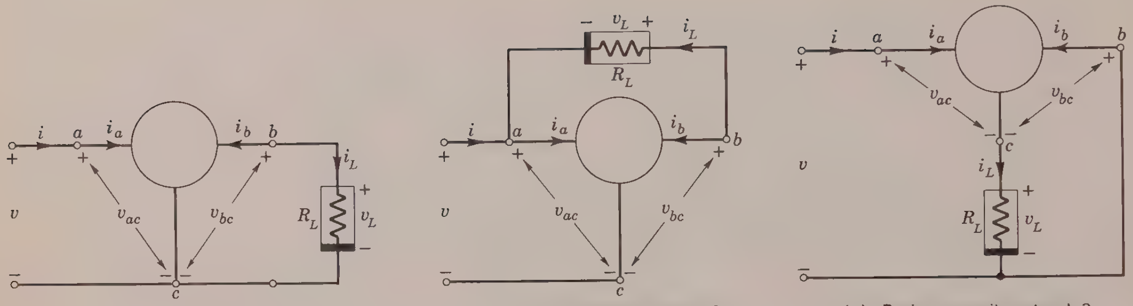
6-7 DP PLOT AND TC PLOT OF NETWORKS CONTAINING THREE-TERMINAL RESISTORS

So far, we have restricted ourselves to series-parallel networks. We shall now consider an important class of networks containing both three-terminal and two-terminal resistors. We shall first develop some graphical techniques for determining the DP plot and the TC plot of the three basic composite networks shown in Fig. 6-38*a* to *c*.

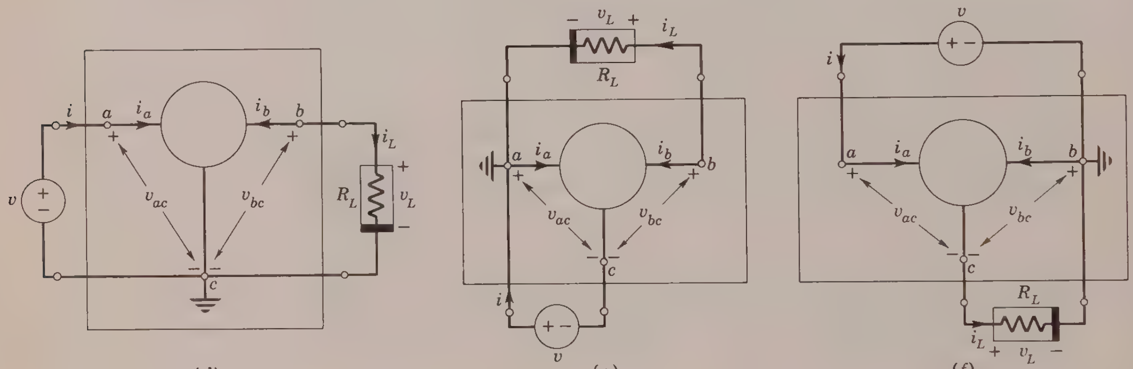
To find the DP plot across terminals *a-c* of basic composite network 1, we connect a voltage source as shown in Fig. 6-38*d* and observe that the resulting network is in the form of basic configuration 2 shown earlier in Fig. 6-7.[1] In this case, the two-terminal resistor on the left is simply a voltage source. Hence, for each value of v we can apply the double load-line method presented earlier to determine the corresponding operating point. From the coordinates of the operating point, we locate a point corresponding to the desired DP plot or TC plot. A repetition of this procedure for different values of the driving-point voltage will therefore allow us to obtain enough points to form the complete DP plot or TC plot.

---

[1] One could alternately choose to drive the circuit with a current source; in this case, the input load lines consist of a family of horizontal lines.



(*a*) Basic composite network 1 (*b*) Basic composite network 2 (*c*) Basic composite network 3



(*d*) (*e*) (*f*)

Fig. 6-38. Three basic composite networks which can be redrawn into the form of basic configuration 2.

Actually, since each input load line is a vertical line, one could draw a family of vertical lines and obtain the DP plot or TC plot by one set of graphical constructions.

Consider next the basic composite networks 2 and 3 in Fig. 6-38*b* and *c*. Each of these networks can be redrawn into the form of basic configuration 2 as shown in Fig. 6-38*e* and *f*, respectively. Hence, the double load-line method is also applicable here. However, since terminals *a* and *b* are now the common terminal, the characteristic curves of the three-terminal resistor must be given with respect to terminal *a* for basic composite network 2 and with respect to terminal *b* for basic composite network 3. Otherwise, the necessary transformation must be made before the double load-line method can be applied. (See Sec. 3-2-3.) This preliminary step is sometimes difficult to carry out in practice because the characteristic curves are seldom given in the entire plane, thereby making both interpolation and extrapolation necessary. We shall now describe another graphical technique which allows us to work directly with the original characteristic curves. Since the principles are identical, we shall consider only one case.

6-7-1 THE TEMPLATE–DOUBLE LOAD-LINE METHOD

Let the three-terminal resistor of the basic composite network 2 be characterized by the same curves shown in Fig. 6-8*a* and *b* and let the v_L-i_L curve of resistor R_L be characterized by the curve shown in Fig. 6-8*d*. For convenience, these curves are redrawn in Fig. 6-39*a* to *c*. The equations of motion for the basic composite network 2 are given by the following:

Equations from the laws of elements

Three-terminal resistor:
$$\begin{cases} i_a = g_1(v_{ac}, i_b) & \text{(6-9}a\text{)} \\ i_b = g_2(v_{bc}, v_{ac}) & \text{(6-9}b\text{)} \end{cases}$$

Resistor R_L:
$$i_L = g(v_L) \qquad \text{(6-9}c\text{)}$$

Equations from the laws of interconnection

$$v = v_{ac} \qquad \text{(6-9}d\text{)}$$

$$i_b = -i_L \qquad \text{(6-9}e\text{)}$$

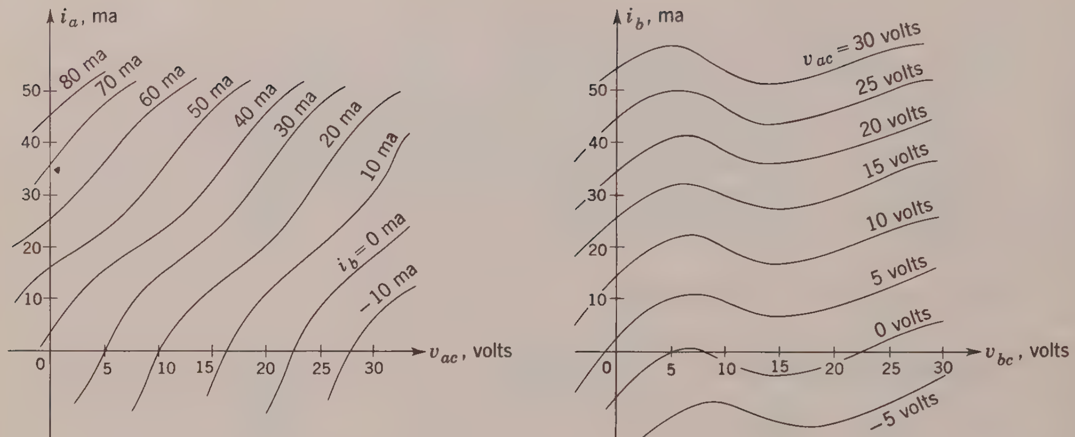
$$i = i_a - i_L = i_a + i_b \qquad \text{(6-9}f\text{)}$$

$$v_L = v_{bc} - v_{ac} \qquad \text{(6-9}g\text{)}$$

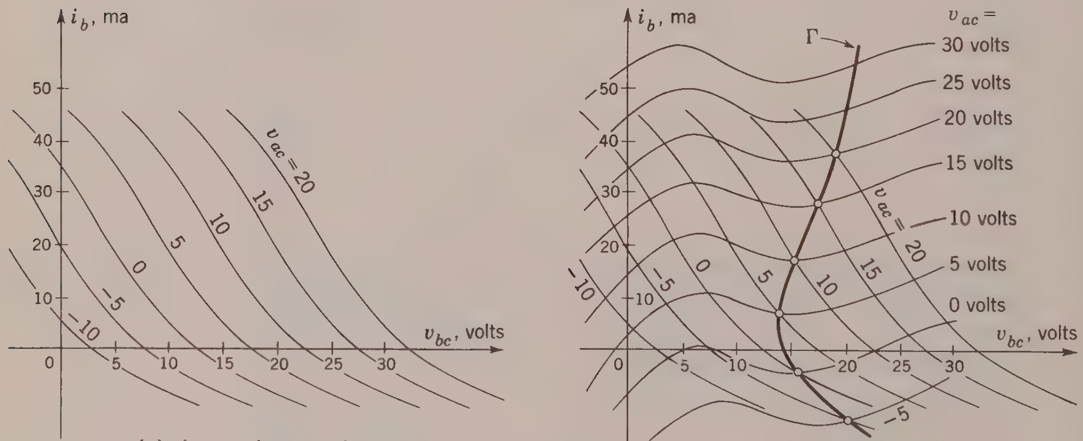
As usual, our procedure consists of eliminating the unwanted variables until a relationship v versus i is obtained for the DP plot. In the case of TC plots, we seek a relationship v_L versus v or v_{bc} versus v. The variables i_L and v_L can be eliminated by substituting Eqs. (6-9*e*) and (*g*) into Eq. (6-9*c*):

$$i_b = -g(v_{bc} - v_{ac}) \qquad \text{(6-10)}$$

A comparison between Eqs. (6-9*b*) and (6-10) shows that they involve the same variables i_b, v_{bc}, and v_{ac}. Hence by superimposing the curves representing these equations, we can obtain a relationship between any two of the three variables. The curves representing Eq. (6-9*b*) are already given by Fig. 6-39*b*. The curves representing Eq. (6-10) can be easily constructed by the following observations. If $v_{ac} = 0$, Eq. (6-10) reduces to $i_b = -g(v_{bc})$ and is simply the load line obtained by reflecting the v_L-i_L curve in Fig. 6-39*c* with respect to the v_L axis, as shown in Fig. 6-39*d*. Next, let us assume a value $v_{ac} = k$ and obtain $i_b = -g(v_{bc} - k)$. The next crucial step consists of the observation that this equation represents a translation of the curve $i_b = -g(v_{bc})$ in Fig. 6-39*d* by k units to the right. Hence, Eq. (6-10) is represented by a family



(a) $i_a = g_1(v_{ac}, i_b)$ (b) $i_b = g_2(v_{bc}, v_{ac})$

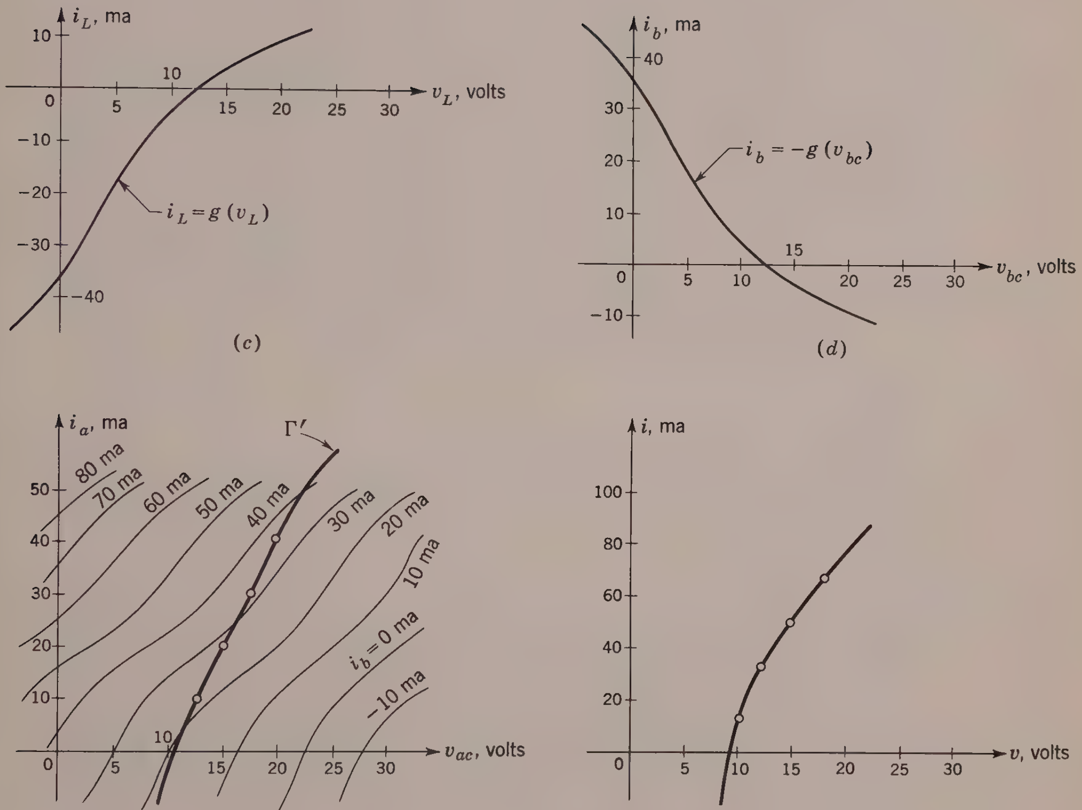


(e) $i_b = -g(v_{bc} - v_{ac})$ (f)

of curves as shown in Fig. 6-39*e* with v_{ac} as a parameter. This can be easily drawn with the help of a template.

We can now superimpose the curves shown in Fig. 6-39*b* and *e* as shown in Fig. 6-39*f*. The intersections between corresponding curves (having the same parameter value v_{ac}) are then identified, and a smooth curve Γ is drawn through these points, as shown in Fig. 6-39*f*.

If we transfer each point on the curve Γ with coordinates (v_{ac}, i_b) into the input-characteristic curves shown in Fig. 6-39*a*, we obtain the corresponding curve Γ′ shown in Fig. 6-39*g*. The value of v_{ac}, i_a, and i_b of each point on Γ′ can now be tabulated. If the DP plot is desired, we make use of Eqs. (6-9*d*) and (*f*) and obtain, corresponding to each point *P* (with $v = v_{ac}$), the current $i = i_a + i_b$.

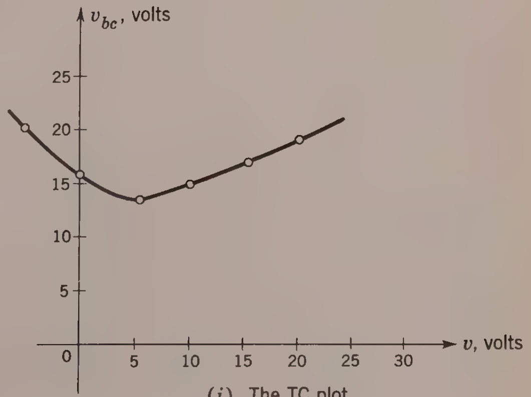


(c)

(d)

(g)

(h) The DP plot



(i) The TC plot

Fig. 6-39. The step-by-step graphical procedure for determining the DP plot and TC plot of composite network 2.

This DP plot is shown in Fig. 6-39*h*. Similarly, if the v_{bc}-vs.-v TC plot is desired, we can obtain the coordinates (v_{ac},v_{bc}) of each point on Γ and obtain the TC plot shown in Fig. 6-39*i*. It is clear that any other relationship, such as the power-transfer plot, can be easily obtained from the coordinates of Γ and Γ′. A review of the above steps shows that both the template method and the double load-line method are used. Accordingly, we shall refer to this method as the *template–double load-line method.* Observe that this method is valid only if the output-characteristic curves are given in the conductance representation, for otherwise, the parameters of the two sets of curves in Fig. 6-39*f* will not be identical. It is therefore necessary that the output-characteristic curves be recast into the conductance representation before the template–double load-line method can be applied to basic composite network 2.

6-7-2 SOME PRACTICAL CIRCUITS

Let us now apply the above techniques to some practical circuits.

EXAMPLES

1. The unijunction transistor circuit shown in Fig. 6-40*a* is called a *relaxation oscillator* and is used in many switching and control applications. Since it is a dynamic network, we are not yet in a position to analyze it. However, as will be shown in Chap. 17, the analysis of this circuit requires that we find the DP plot across the capacitor terminals. This DP plot can be found by observing that it is simply the parallel combination of the 2 two-terminal black boxes shown in Fig. 6-40*b*, where the black box *N* represents the unijunction transistor circuit shown in Fig. 6-40*c*. Our first step, therefore, is to find the DP plot across *N*.

   The characteristic curves of the unijunction transistor type 2NXX21 are given in Appendix D. If we draw the output load line on top of the output-characteristic curves as shown in Fig. 6-40*d*, we obtain a set of points α, β, γ, δ, ε, ξ, etc. The coordinates (V_{BB},I_E) of these points are plotted on top of the input-characteristic curves as shown in Fig. 6-40*e*. The solid curve drawn through these points is the DP plot across the black box *N*.

   To obtain the DP plot across the capacitor terminals, we apply the parallel-combination techniques to the circuit in Fig.

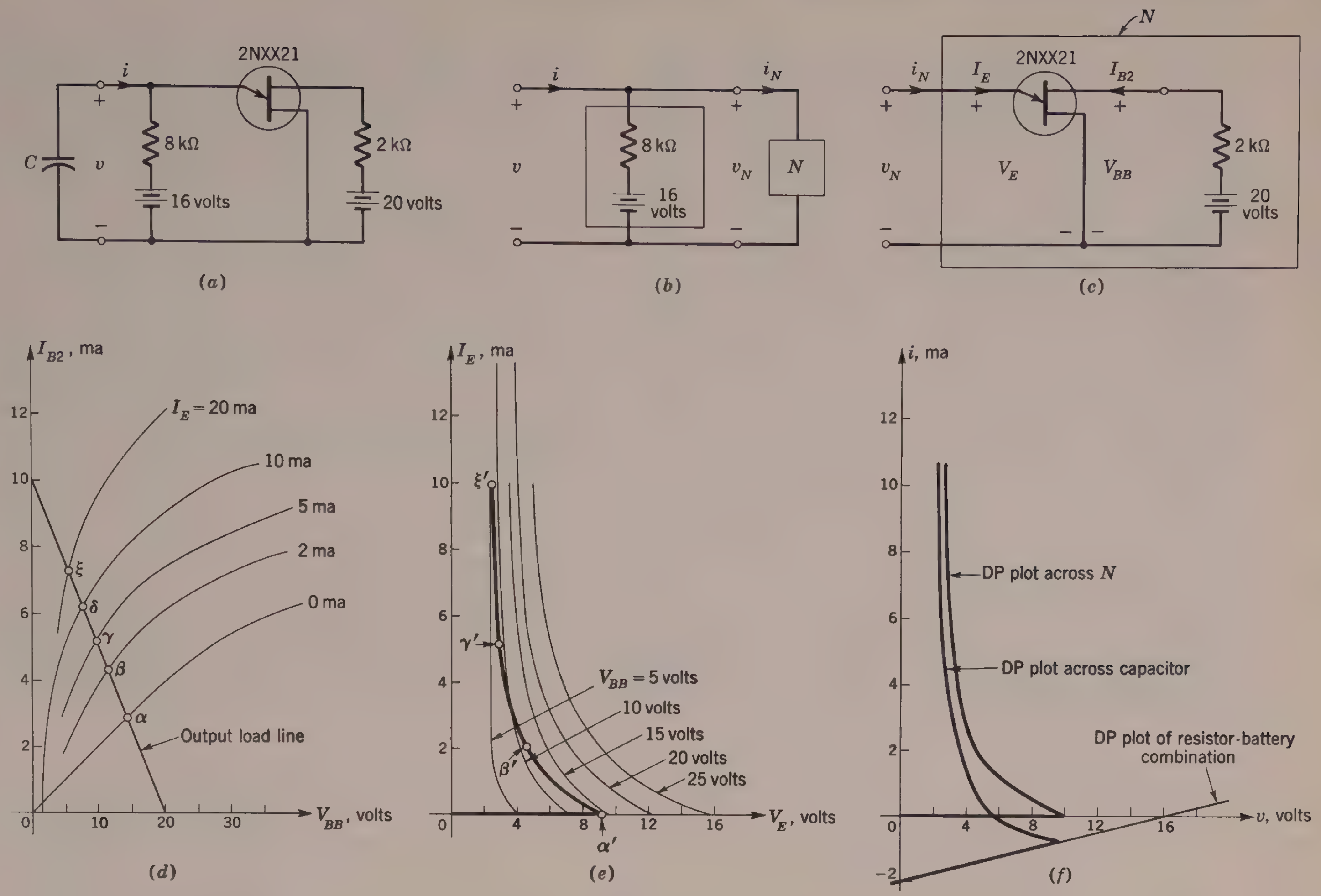


Fig. 6-40. The graphical procedure for determining the DP plot across the capacitor terminals of a unijunction transistor relaxation oscillator.

6-40*b* and obtain the desired i-vs.-v DP plot. It will be shown in Chap. 17 that if a capacitor is connected across the composite network, a periodic sawtooth waveform will be generated. This waveform is, in turn, used in many applications such as oscilloscopes and television sets.

2. The FET circuit shown in Fig. 6-41*a* is useful in many digital computer applications.[1] In this case, it is the TC plot between v_o and v_i that is important. Since the output terminal does not draw any current, the circuit can be redrawn into the form of basic composite network 1 as shown in Fig. 6-41*b*, where the black box N represents the upper FET shown in Fig. 6-41*c*.

   We must first find the i-vs.-v DP plot across the black box N. This is simply obtained by first rotating the output-characteristic curve (see Appendix D) corresponding to $v_{GS} = 0$ by 180° (since the source terminal is assigned the positive reference) and then translating the curve by 20 volts as shown in Fig. 6-41*d*. The load line corresponding to this DP plot is then drawn on top of the output-characteristic curves of the lower FET. The resulting TC plot is obtained as shown in Fig. 6-41*f*. The reader should now recognize the usefulness of this TC plot. Among other applications, it can be used to transform a sinusoidal input voltage into a square wave.

3. As a final example, consider the tunnel diode-transistor circuit shown in Fig. 6-42*a*. This circuit is used in practice to increase the current level of the tunnel diode, an inherently low-power device. To be specific, let the tunnel diode v_d-i_d curve be as shown in Fig. 6-42*b*. Since this circuit corresponds to the basic composite network 2 in Fig. 6-38*b*, we can apply the template–double load-line method. The family of i_b-vs.-v_{bc} curves is shown in Fig. 6-42*c*.[2] If we represent the base-to-emitter characteristics of the *n-p-n* transistor by a single curve as shown in Fig. 6-42*d*, then the intersections between this curve and the above family of curves yield a set of points with coordinates (v_{ac}, i_b). These points can be transferred into the collector-to-emitter characteristic curves of the transistor as shown in Fig. 6-42*e*. Since $i = i_d + I_C = i_B + I_C \approx I_C$, the resulting solid curve is the desired DP plot. Notice that this curve resembles that of a tunnel diode but with one significant difference; namely, the current ordinate is approximately 100 times larger than the original v_d-i_d curve shown in Fig. 6-42*b*. In other words, we have transformed a low-current tunnel diode into a high-current voltage-controlled nonlinear resistor whose v-i curve resembles

---

[1] It is a common practice to omit drawing the voltage sources as well as output voltage terminals which are measured with respect to ground.

[2] Notice that the reference polarity and direction of the tunnel diode must first be redefined to agree with the network shown in Fig. 6-38*b*.

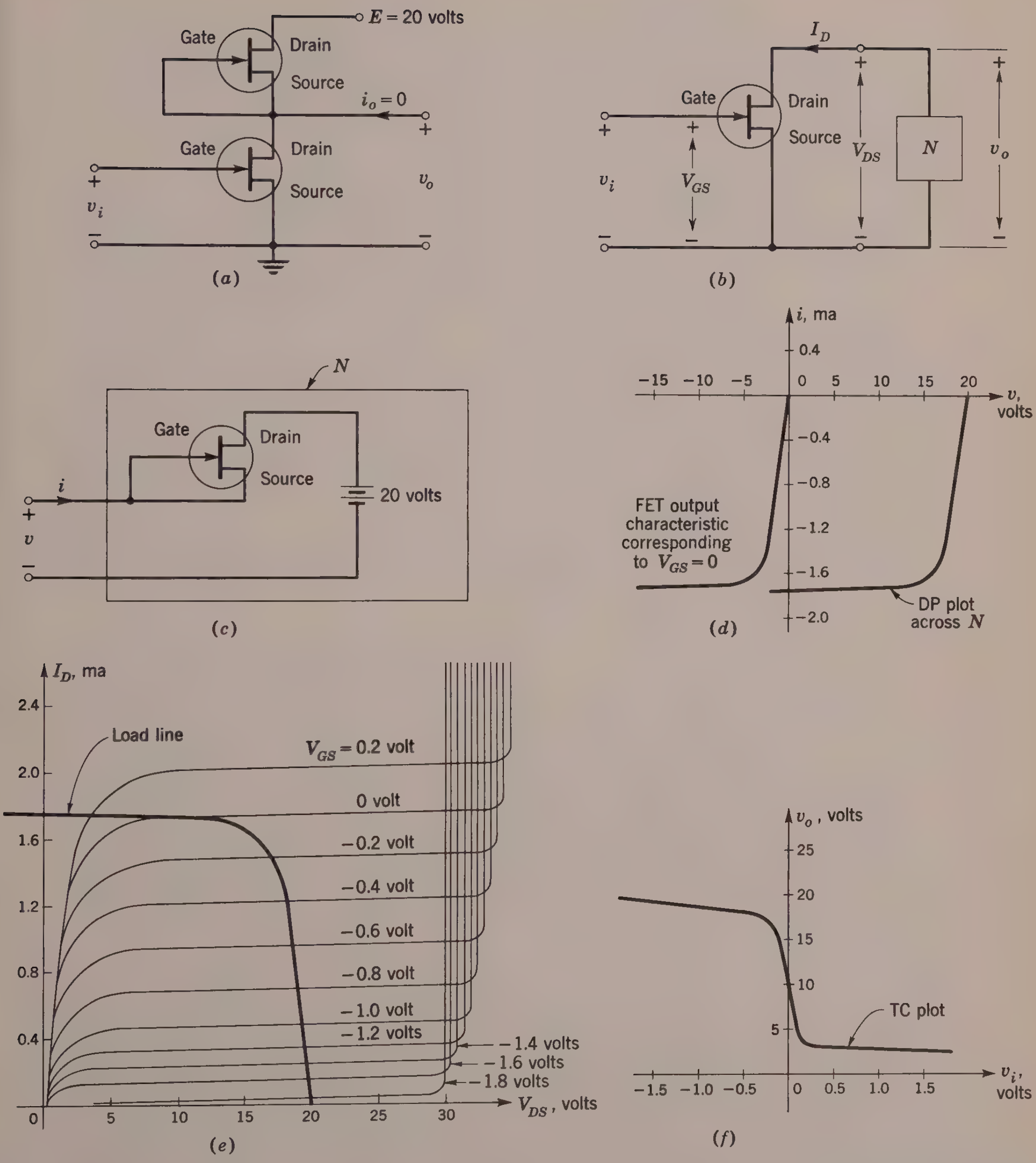


Fig. 6-41. The graphical procedure for determining the TC plot of an FET logic circuit.

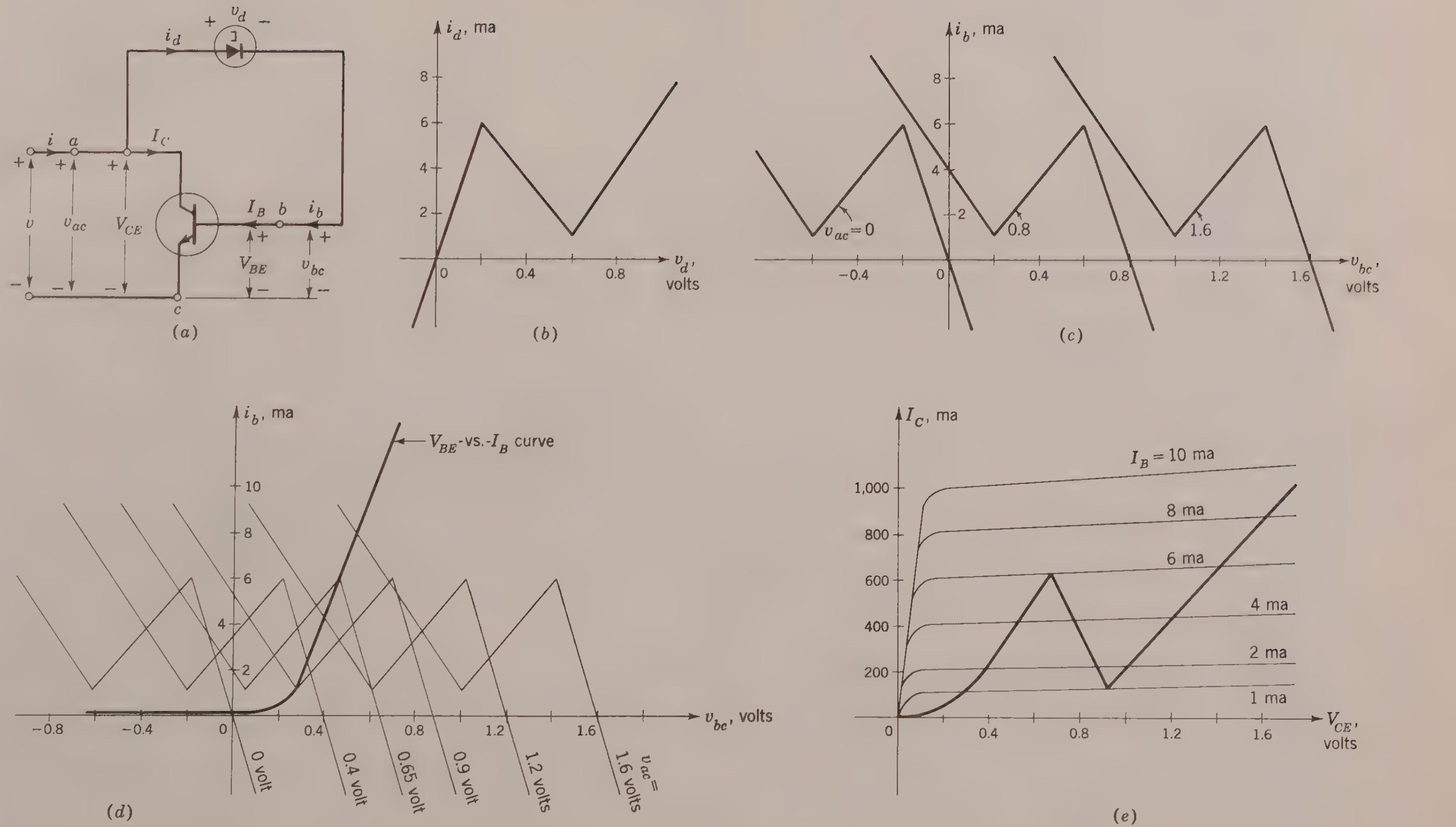


Fig. 6-42. The graphical procedure for determining the DP plot of a tunnel diode-transistor composite circuit.

that of a tunnel diode. Hence, this circuit can be considered to be a rough approximation of the current scalor introduced earlier.[1] The advantage of the present circuit, however, is that it does not require an external power supply.

Exercise 1: Describe the template–double load-line method for determining the DP plot and TC plot of the basic composite network 3 shown in Fig. 6-38*c*. HINT: Assume the output-characteristic curves in the resistance representation.

Exercise 2: In many practical circuits, the resistor R_L in Fig. 6-38 consists simply of a linear resistor in series with a battery. Give a simplified procedure for determining the DP plot and TC plot for each of the basic composite networks shown in Fig. 6-38.

Exercise 3: Find the v_{BB}-vs.-v_N TC plot of the unijunction transistor circuit shown in Fig. 6-40*c*.

Exercise 4: (*a*) Find the TC plot of the FET logic circuit shown in Fig. 6-41*a* but with the FET replaced by a *p*-channel type 2NXX20, and with $E = -20$ volts. (*b*) Sketch the output voltage corresponding to a 10-volt peak-to-peak sinusoidal input voltage.

Exercise 5: Find the DP plot of the tunnel diode-transistor circuit shown in Fig. 6-42*a* but with the diode terminals interchanged and with the transistor replaced by a *p-n-p* transistor.

6-8 COMPOSITE CHARACTERISTICS OF THREE-TERMINAL RESISTIVE BLACK BOXES

In order to apply the graphical techniques to more complex resistive networks, it is often necessary to isolate appropriate three-terminal black boxes and replace them by three-terminal resistors having the same sets of external-characteristic curves. The procedure for obtaining these external-characteristic curves is based on the successive application of graphical techniques for analyzing the four basic three-terminal black boxes shown in Fig. 6-43.

6-8-1 FOUR BASIC THREE-TERMINAL BLACK BOXES

Consider first the black box N shown in Fig. 6-43*a*. To obtain the input-characteristic curves, we connect a voltage source in series with resistor R_2 and determine the DP plot across terminals *a-c* corresponding to each parameter value v_2. This can be done by observing that the DP plot is simply the series combination of R_1 and a two-terminal black box N_2 in the form of basic composite network 1 (Fig. 6-38*a*). The output-characteristic curves are obtained in an exactly similar manner; that is, a voltage source

---

[1] The reader will recall that the current scalor introduced in Sec. 3-7-1 is intended for the same purpose while preserving the exact form of the *v-i* curve.

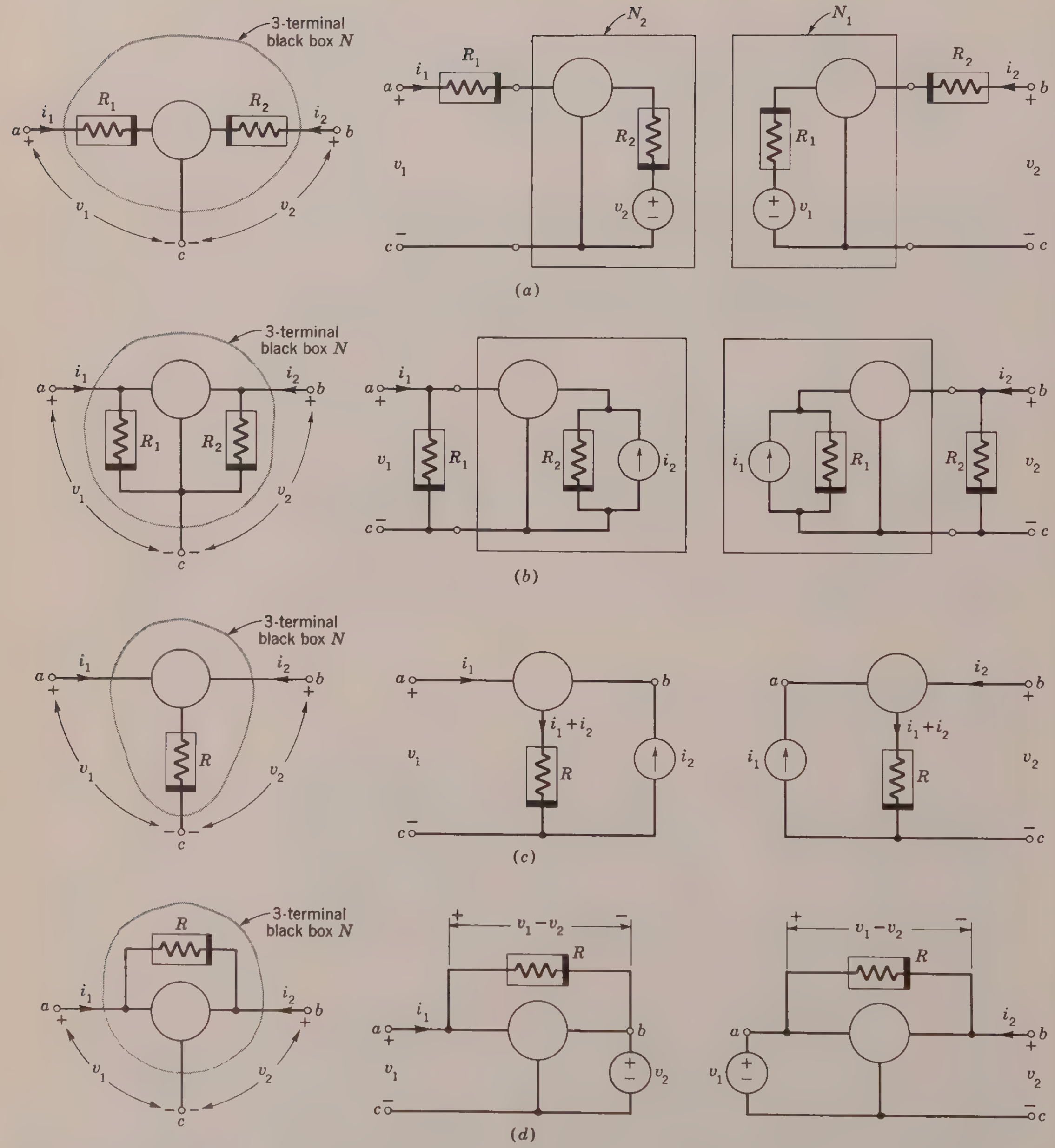


Fig. 6-43. Four basic three-terminal black boxes and the appropriate circuit arrangements for obtaining their DP plots.

is connected in series with R_1, and the DP plot across terminals b-c is determined corresponding to each parameter value v_1. Notice that the three-terminal resistor inside the black box may be characterized by either the conductance representation or the resistance representation. The resulting characteristics, however, will be in the conductance representation because the parameters are chosen to be v_2 and v_1.

The three-terminal black box N shown in Fig. 6-43*b* is, roughly speaking, the dual of that shown in Fig. 6-43*a*. Hence, the parameters are now chosen to be i_2 and i_1, and the characteristic curves may be obtained by a dual procedure. In this case, the characteristic curves are obtained in the resistance representation.

The easiest method for obtaining the input-characteristic curves of the three-terminal black box N shown in Fig. 6-43*c* is to connect a current source across terminals *b-c* and determine the DP plot across terminals *a-c*. This is done by observing that the resulting two-terminal black box is equivalent to two resistors in series. The *v-i* curve of the first resistor is simply the characteristic curve of the three-terminal resistor corresponding to $i_2 = k$. The *v-i* curve of the second resistor is obtained by translating the *v-i* curve of the resistor R by $i_2 = k$ units along the current axis i_1.[1] A similar procedure can be used to obtain the output-characteristic curves. Notice that in this case, the characteristic curves of the three-terminal resistor must be given in the resistance representation.[2]

The three-terminal black box N shown in Fig. 6-43*d* is, roughly speaking, the dual of that shown in Fig. 6-43*c*. Instead of the current sources i_2 and i_1, we now connect the voltage sources v_2 and v_1. The input-characteristic curve corresponding to each parameter value $v_2 = k$ is simply the DP plot of two resistors in parallel. The *v-i* curve of the first resistor is the characteristic curve of the three-terminal resistor corresponding to $v_2 = k$. The *v-i* curve of the second resistor is obtained by translating the *v-i* curve of the resistor R by $v_2 = k$ units along the voltage axis v_1. The same procedure can be used to obtain the output-characteristic curves. Notice that in this case, the characteristic curves of the three-terminal resistor must be given in the conductance representation.

6-8-2 CASCADE CONNECTION OF BASIC THREE-TERMINAL BLACK BOXES

By an appropriate combination of elements into black boxes, it is possible to determine the DP plots and TC plots of rather complex, nonlinear networks containing both two-terminal and three-terminal resistors. We shall illustrate the basic principle with the help of the network shown in Fig. 6-44*a*.

The first step is to combine N_2 and R_{12} into a three-terminal

---

[1] This translation is necessary because the current entering the resistor R in Fig. 6-43*c* is equal to $i_1 + i_2$.

[2] Notice that if the current sources are replaced by voltage sources, the graphical methods discussed so far are not applicable. In this case, we must appeal to the *v*-shift theorem described in the next chapter.

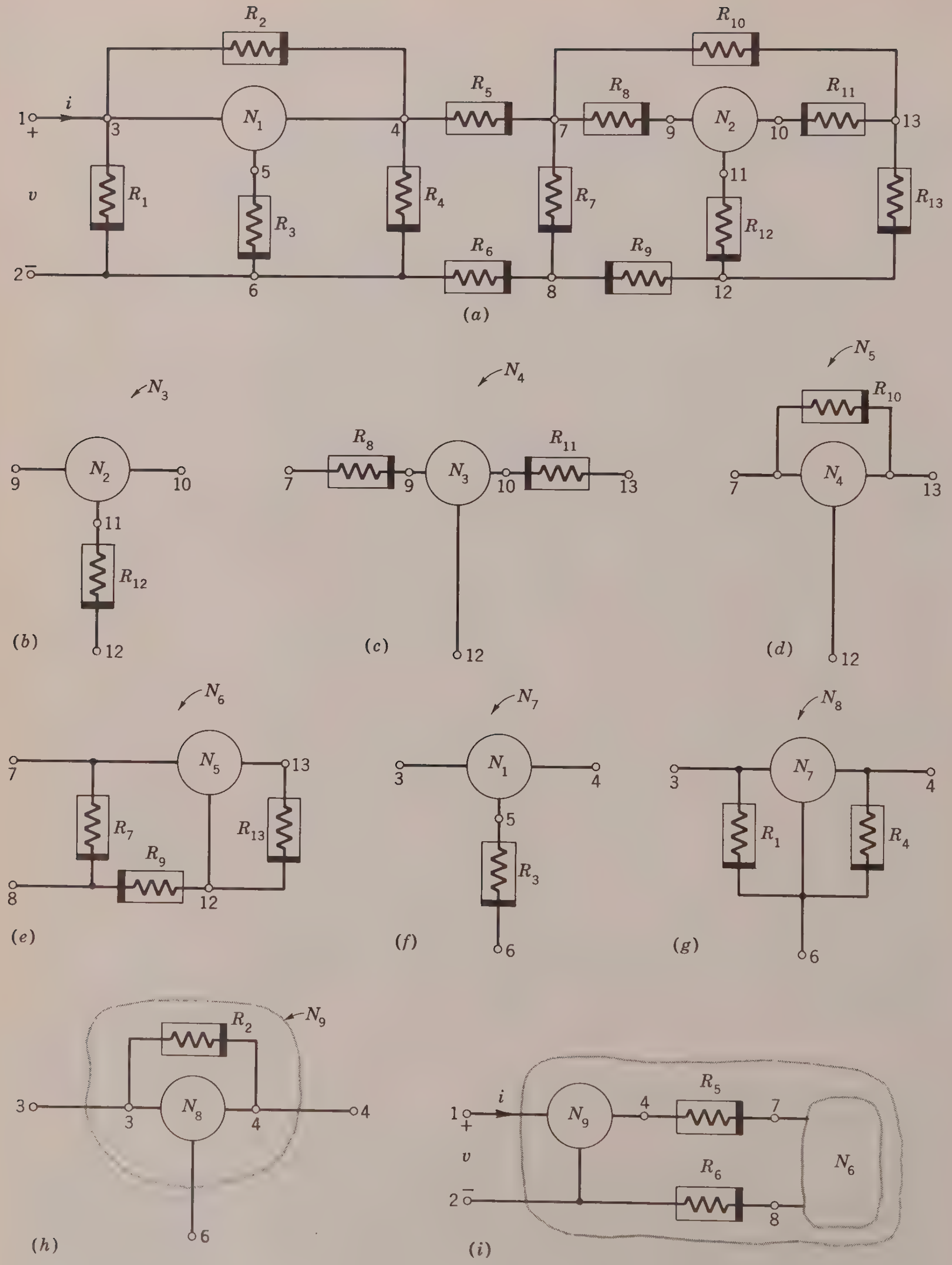


Fig. 6-44. The successive reduction procedure for determining the DP plot by graphical techniques.

black box N_3 as shown in Fig. 6-44*b*. Notice that this corresponds to the black box shown in Fig. 6-43*c*, and hence its characteristic curves can be found by the method described earlier. The next step is to combine N_3 with R_8 and R_{11} to form the new three-terminal black box N_4 shown in Fig. 6-44*c*. This corresponds to the black box shown in Fig. 6-43*a*. Combining next N_4 with R_{10}, we obtain the new black box N_5 as shown in Fig. 6-44*d*. This corresponds to the black box in Fig. 6-43*d*. If we now combine N_5 with R_7, R_9, and R_{13} to obtain the two-terminal black box N_6 shown in Fig. 6-44*e*, we can determine the DP plot across terminals 7 and 8 of this black box. Hence, the entire network to the right of terminals 7 and 8 (including R_7) can be replaced by the two-terminal black box N_6. Notice that it is possible to do this because none of the resistors on the right of terminals 7 and 8 are connected to elements on the left of terminals 7 and 8. The entire black box N_6 is, therefore, said to be *in cascade* with the network on the left of terminals 7 and 8.

We can repeat the same procedure for the remaining network. The steps are indicated in Fig. 6-44*f* to *i*. The final DP plot can be determined by the procedure described in Sec. 6-7.

In general, any network made up of cascade connections of subnetworks can be handled by the same method. For complicated networks, a general computer program can be written to implement the above steps automatically. If the network cannot be broken up in the manner described above, the more general techniques presented in Chaps. 7 and 12 should be used.

Exercise 1: If the voltage sources in Fig. 6-43*a* are replaced by current sources i_2 and i_1, respectively, describe the procedure for determining the characteristic curves of the three-terminal black box if the three-terminal resistor is characterized by: (*a*) the conductance representation; (*b*) the resistance representation. HINT: One of the two resistors is redundant in each case.

Exercise 2: Repeat Exercise 1 for the black box shown in Fig. 6-43*b* but with the current sources replaced by voltage sources v_2 and v_1, respectively.

Exercise 3: Describe the procedure for determining the characteristic curves of the three-terminal black box shown in Fig. 6-43*b* with i_2 and i_1 as the parameter.

Exercise 4: In many practical three-terminal resistors, the input-characteristic curves can be approximately represented by one curve. Show how the procedures described for the four black boxes in Fig. 6-43 can be simplified under this condition.

Exercise 5: The order for carrying out *some* of the steps shown in Fig. 6-44 can be interchanged. Find all such possibilities and describe some alternate methods for determining the DP plot.

6-9 SUMMARY

Graphical methods for determining operating points

1. The load-line method is valid for 2 two-terminal black boxes in parallel (basic network configuration 1).
2. The double load-line method is valid for the case where a three-terminal resistor is terminated by a two-terminal black box across both ends (basic network configuration 2).

Graphical methods for determining DP plots

1. The series-combination technique for two resistors connected back to front in series.
2. The parallel-combination technique for two resistors connected back to back in parallel.

Graphical methods for determining TC plots

1. The template method for a nonlinear voltage divider or a nonlinear current divider.
2. The graphical elimination method for a nonlinear voltage divider or a nonlinear current divider.

The template–double load-line method This method is useful for determining the DP plots and TC plots of networks containing both three-terminal and two-terminal resistors which can be reduced into a combination of three basic composite networks. This method can be extended to determine the characteristic curves of three-terminal black boxes made up of both three-terminal and two-terminal resistors. The possibility for doing this allows us to determine the DP plots and TC plots of resistive nonlinear networks which can be broken up into cascade connections of simpler networks amenable to the preceding methods.

PROBLEMS

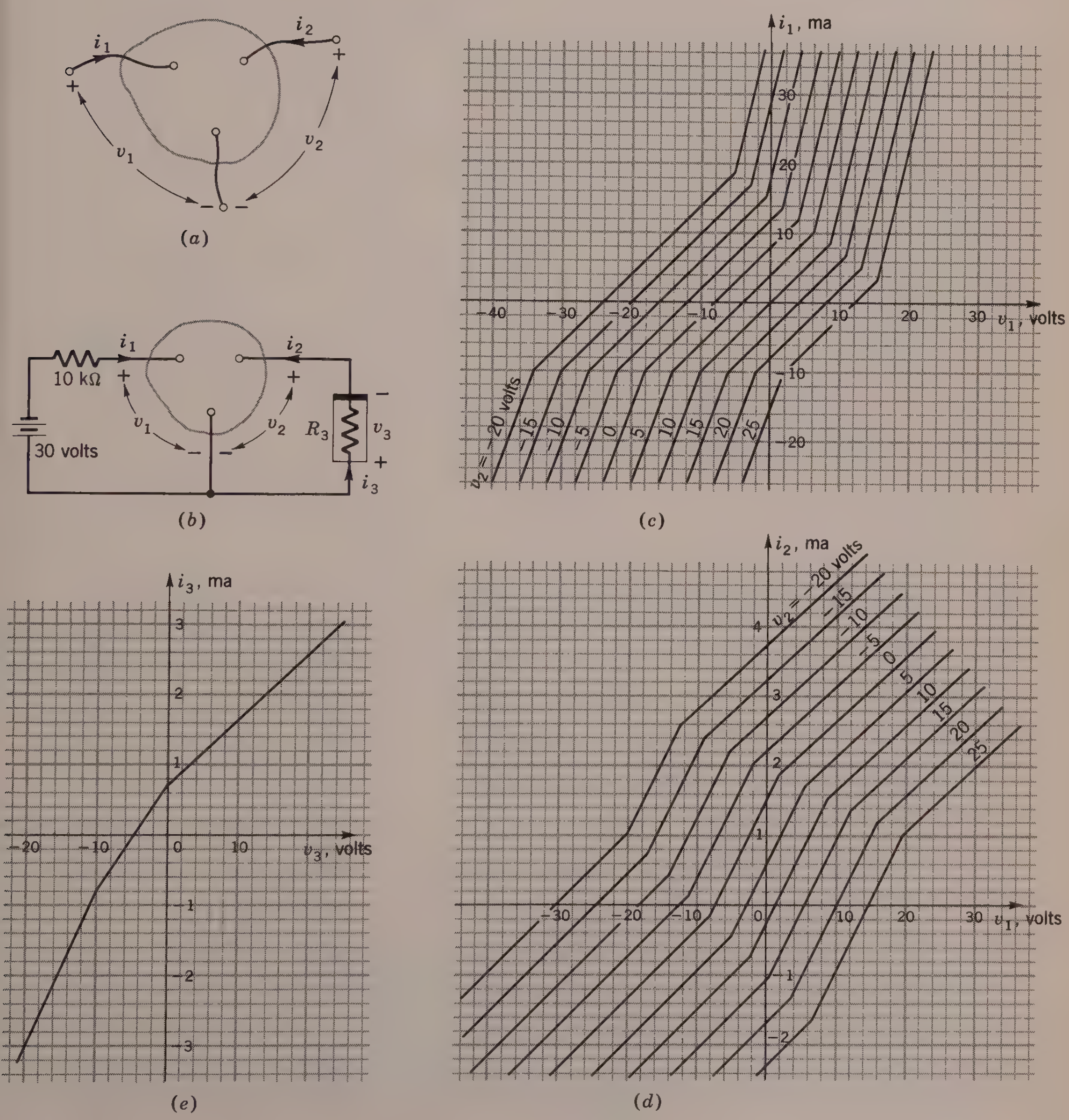
6-1 A piece of rock picked up from the moon was analyzed for its electrical properties by attaching three electrodes to the rock's surface, forming an equilateral triangle as shown in Fig. P6-1*a*. A series of measurements was obtained, and the data were plotted as the two families of curves shown in Fig. P6-1*c* and *d*.

(a) Is this three-terminal black box passive or active? Explain why.

(b) What type of basic element does this rock represent?

(c) If the rock is connected as shown in Fig. P6-1b, find the operating point $(i_1,i_2,i_3,v_1,v_2,v_3)$. The v-i curve for nonlinear resistor R_3 is given in Fig. P6-1e.

Fig. P6-1.



(a) (b) (c) (e) (d)

6-2 Find the DP plots for the networks shown in Fig. P6-2*a* to *j*. Use the piecewise-linear *v-i* curves for the zener diode and the constant-current diode shown, respectively, in Figs. 6-20*c* and 6-21*b* in the text.

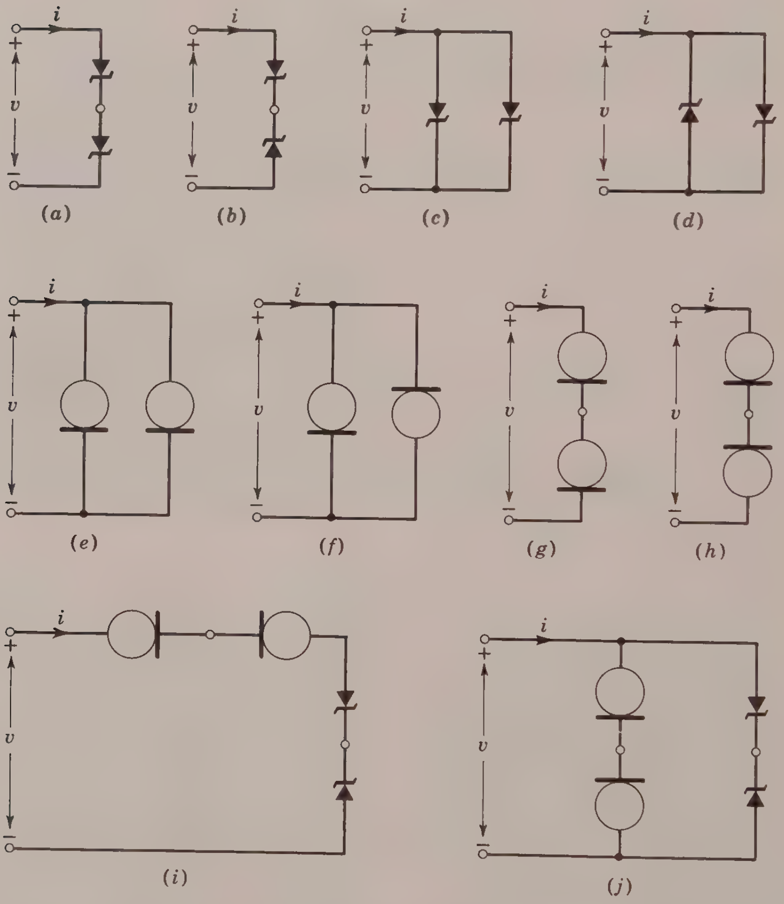


Fig. P6-2.

6-3 Find the DP plots for the ideal diode networks shown in Fig. P6-3*a* to *h*.

6-4 Prove the following statements and give an example in each case.

(a) The DP plot across two or more current-controlled resistors in series is also current-controlled.

(b) The DP plot across two or more voltage-controlled resistors in parallel is also voltage-controlled.

(c) The DP plot across any series-parallel network containing only monotonically increasing resistors is also monotonically increasing.

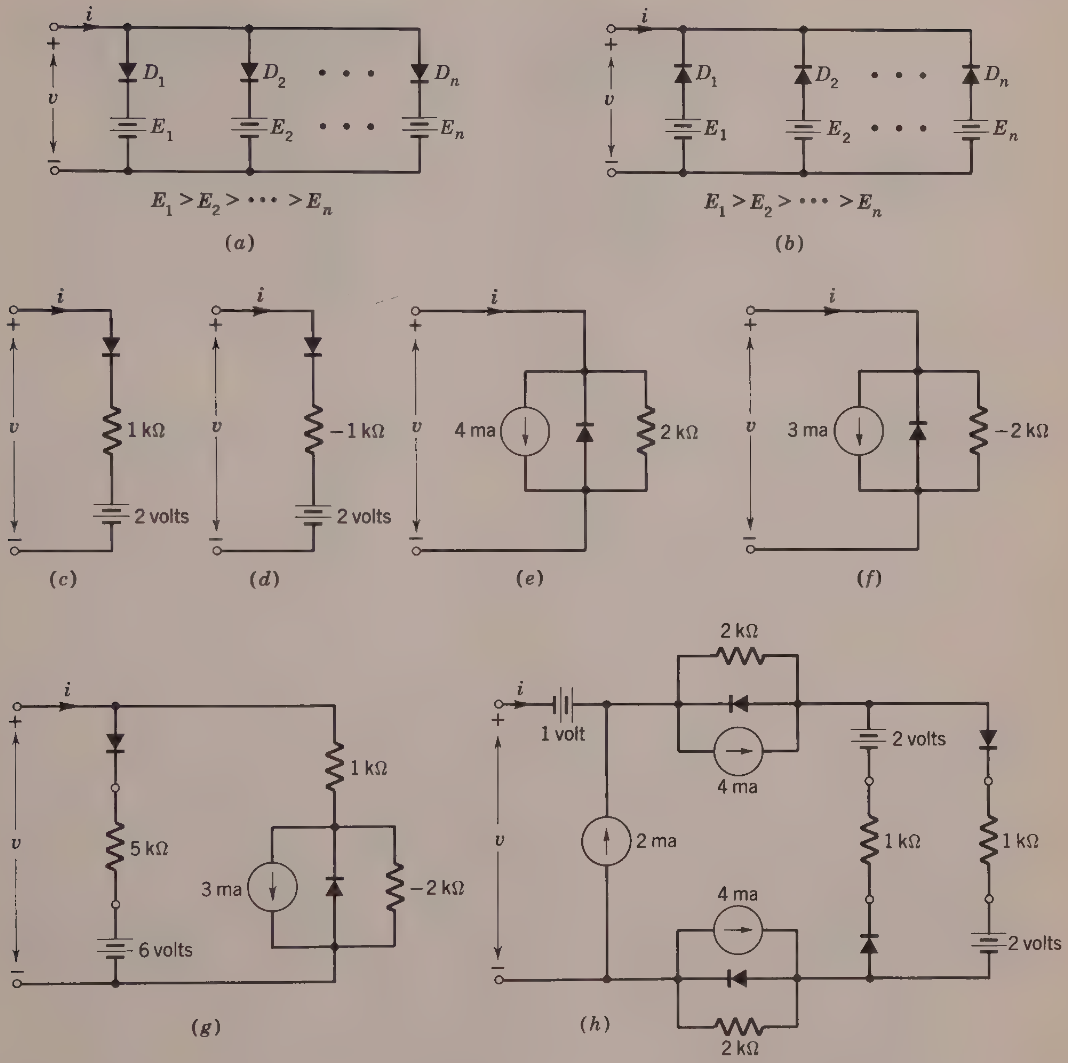


Fig. P6-3.

(d) The DP plot across two identical current-controlled resistors in series is obtained by multiplying the abscissa of the original v-i curve by 2. Conversely, any current-controlled resistor can be decomposed into two identical resistors in series.

(e) The DP plot across two identical voltage-controlled resistors in parallel is obtained by multiplying the ordinate of the original v-i curve by 2. Conversely, any voltage-controlled resistor can be decomposed into two identical resistors in parallel.

6-5 Find the DP plot for the networks shown in Fig. P6-5*a* to *d*. The *v-i* curves for the resistors R_A and R_B are shown in Fig. P6-5*e* and *f*, respectively.

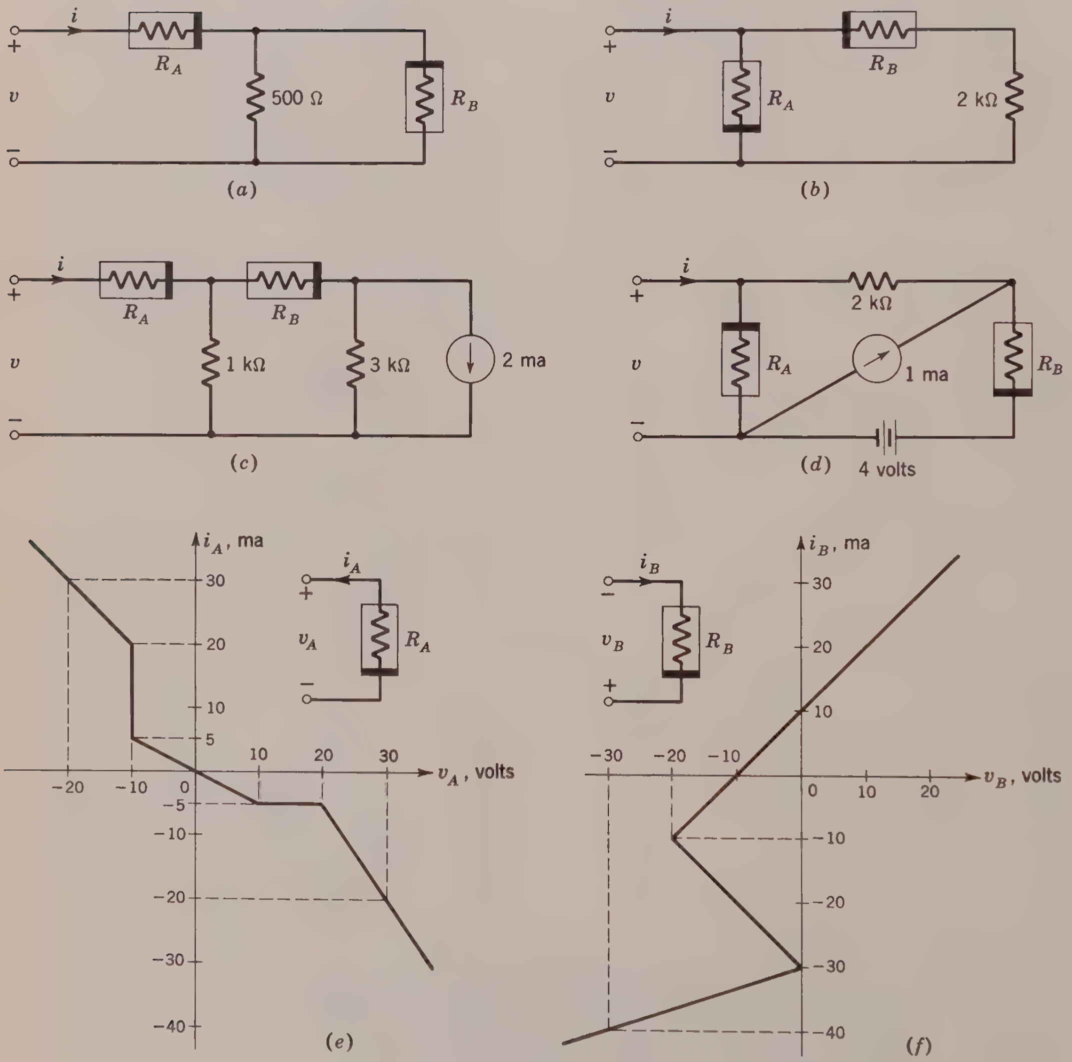


Fig. P6-5.

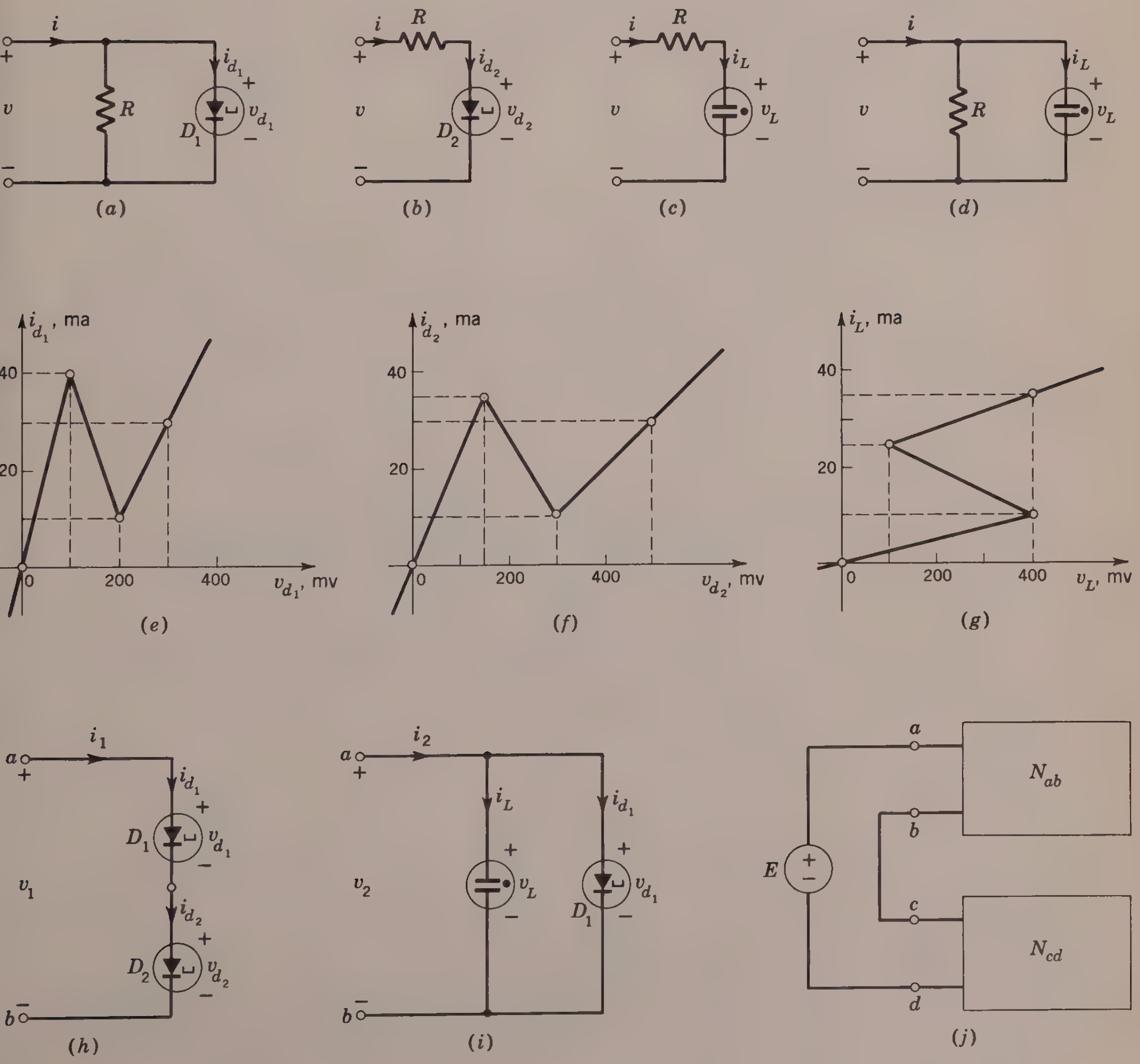
6-6 (a) Depending on the value of the resistance R in the networks shown in Fig. P6-6*a* to *d*, the resulting DP plot may be monotonically increasing, voltage-controlled, current-controlled, or even multivalued. Find the range of R in each network for which the DP plot will assume the above types. The v-i curves of the tunnel diodes and the glow lamp are shown in Fig. P6-6*e* to *g*.

(b) Find the DP plot of the network shown in Fig. P6-6*h*.

(c) Find the DP plot of the network shown in Fig. P6-6*i*.

(d) If the networks in Fig. P6-6*h* and *i* are connected as shown in Fig. P6-6*j*, find the operating points corresponding to $E = 50$, 100, and 500 mv.

Fig. P6-6.



(a) (b) (c) (d) (e) (f) (g) (h) (i) (j)

6-7 A circuit found in an optical instrument consists of an *n-p-n* transistor, a back diode, and two varistors connected as shown in Fig. P6-7.

(a) Find the operating point of this network with $v_1 = 0.9$ volt.

(b) Find the v_4-vs.-v_1 TC plot.

(c) Find the power-transfer plot between each varistor and the driving-point voltage v_1.

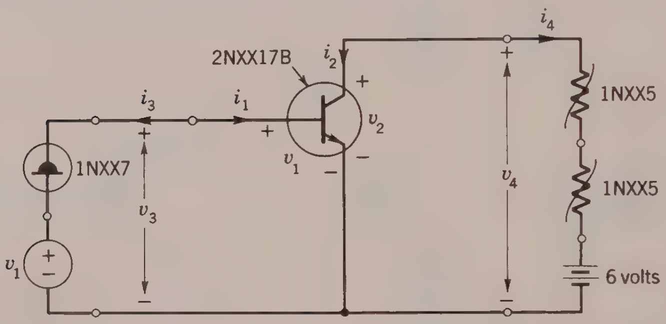


Fig. P6-7.

6-8 A 2NXX17B transistor is connected in the circuit shown in Fig. P6-8. The total power dissipated in the transistor is the sum of the power into the base-emitter terminals ($I_B V_{BE}$) plus the power into the collector-emitter terminals ($I_C V_{CE}$). It is known that the power $I_B V_{BE}$ is much less than the power $I_C V_{CE}$. Thus the total power dissipated by the transistor is very nearly $I_C V_{CE}$.

(a) Find the TC plots v_o versus v and i_C versus v, and then use these TC plots to calculate the power dissipation of the transistor as a function of the driving-point voltage v.

(b) Repeat (*a*) for $R_L = 120\ \Omega$. Assume that the maximum power dissipation of the transistor is 0.25 watt and find the maximum allowable input voltage v.

(c) If the maximum power dissipation of the transistor is 0.25 watt and R_L may vary, find the value of R_L such that the power dissipated in R_L is a maximum without exceeding the transistor power limitations.

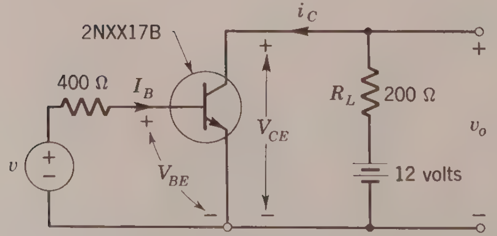
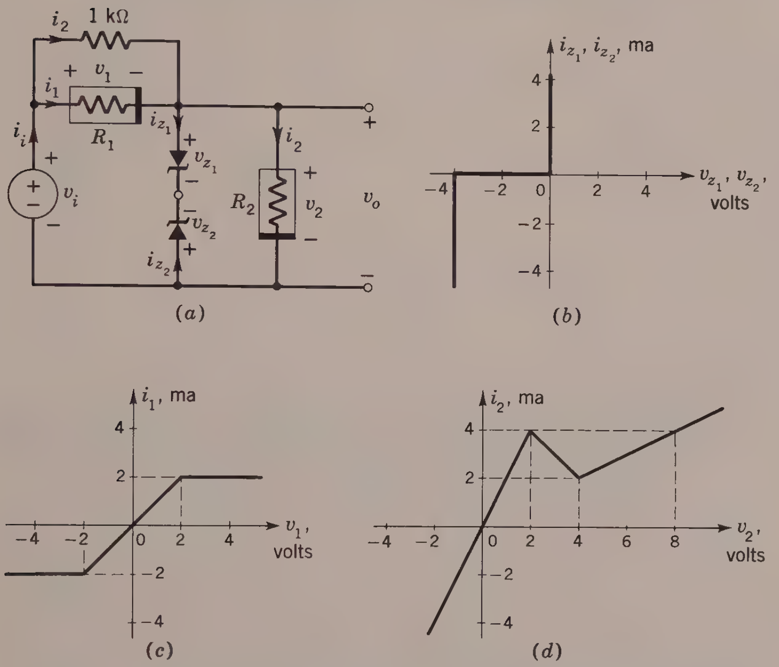


Fig. P6-8.

6-9 The network shown in Fig. P6-9*a* is part of a control circuit for a rocket stabilization system. The zener diodes are characterized by the same v_z-i_z curve shown in Fig. P6-9*b*. The v_1-i_1 and v_2-i_2 curves of R_1 and R_2 are shown in Fig. P6-9*c* and *d*, respectively.

(a) Find the v_o-vs.-v_i TC plot.

(b) Find the power-transfer plot from the resistor R_2 to the driving-point voltage v_i.

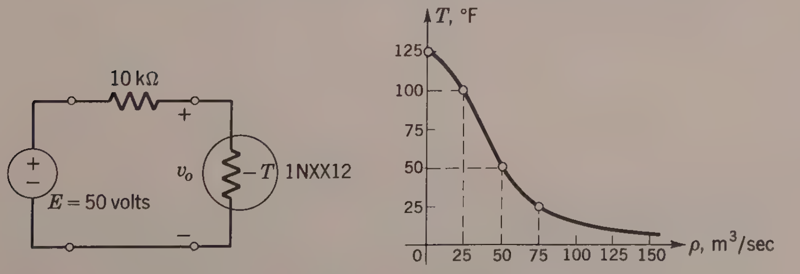


(*a*) (*b*)

(*c*) (*d*)

Fig. P6-9.

6-10 As part of a control system in a petroleum refinery, it is necessary to sense the rate of flow of propane in a pipe. The circuit shown in Fig. P6-10*a* uses an NTC thermistor and can be designed to do the job. Since the rate of cooling is affected by the flow rate of the gas, the ambient temperature T of the thermistor is a function of the flow rate ρ (cubic meter per second). This relationship was obtained by measurement and is shown in Fig. P6-10*b*. Find the relationship between the output voltage v_o and the flow rate ρ.



(*a*) (*b*)

Fig. P6-10.

6-11 The two networks shown in Fig. P6-11 are usually called hybrid circuits because they contain both transistors and tunnel diodes. These circuits are useful in the design of many alarm and overload protection systems. Describe the step-by-step procedure for determining the DP plots of these circuits.

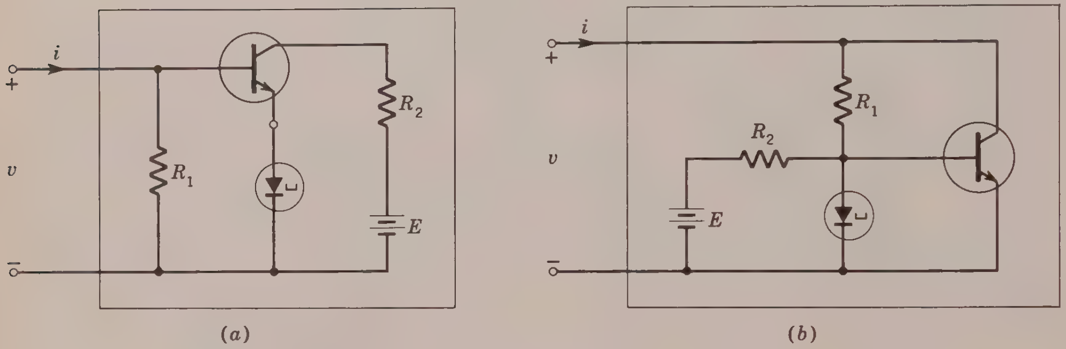


Fig. P6-11.

6-12 The three networks shown in Fig. P6-12 are widely used in practice as voltage regulators. Describe the step-by-step procedure for determining the v_o-vs.-v_i TC plot for each circuit.

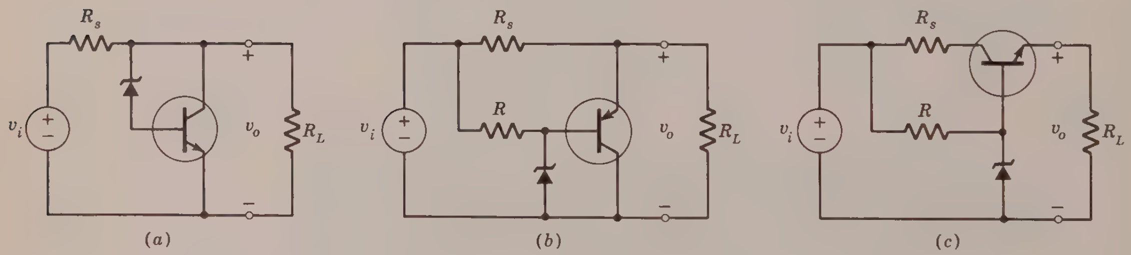


Fig. P6-12.

7 PRINCIPLES OF EQUIVALENCE AND SYMMETRY

7-1 SOME SHORTCUTS IN NONLINEAR NETWORK ANALYSIS

The graphical approach presented in the preceding chapter has two significant advantages over the piecewise-linear approach (Chap. 12) and the numerical techniques. First, it provides the engineer with a lot of qualitative information in addition to the network solution. For example, the effect of the nonlinearity of the resistor *v-i* curves on the final DP plot is easily established. This information is extremely important in any practical design for which the effect of temperature changes, component aging, etc., on the network's performance must be determined. Second, it provides all possible solutions in one shot. This is because the amount of work required by the graphical approach depends primarily on the network topology, not on the type of *v-i* curves.[1]

The most serious disadvantage of the graphical approach is that it lacks generality. There are many circuits, such as non-series-parallel networks, which cannot be analyzed by the methods described in Chap. 6. Our object in this chapter is to present several techniques which will allow us to analyze a much larger class of networks. These techniques generally consist of some shortcuts which permit a complex network to be transformed into a simpler, but equivalent, network. These shortcuts are based on the principles of equivalence and symmetry.

7-2 DEFINITION OF EQUIVALENT NETWORKS

The concept of replacing a network or part of a network by a black box as presented in Chap. 6 is actually an application of the general concept of equivalent networks. Intuitively, it seems reasonable to say that two *n*-terminal black boxes, N_1 and N_2, are

---

[1] In sharp contrast with this, the amount of work required by the piecewise-linear method depends primarily on the type of *v-i* curves, not on the network topology. With the current state of the art, the numerical techniques can not be used to analyze networks with more than one solution.

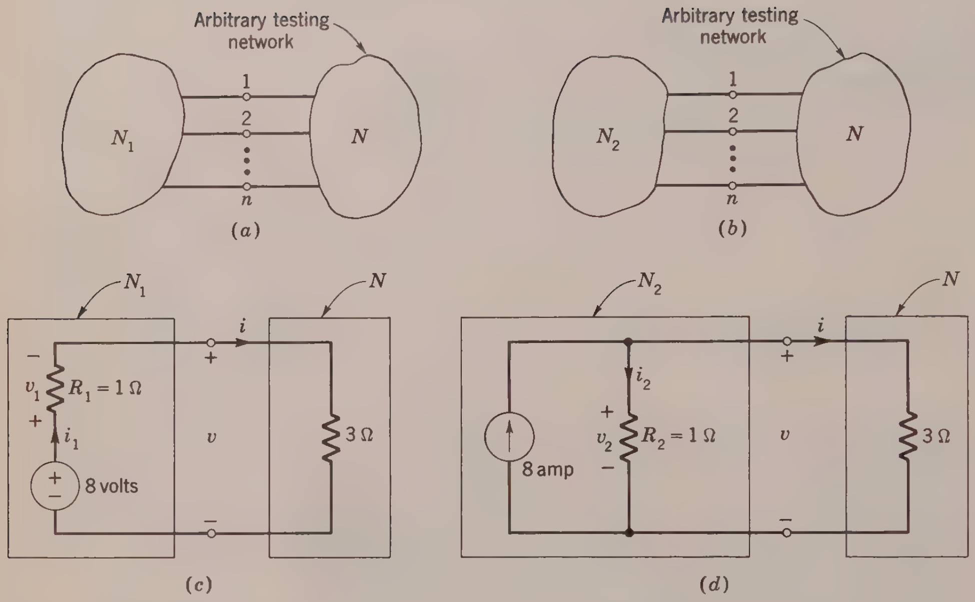


Fig. 7-1. An experiment for testing the equivalence between two black boxes N_1 and N_2.

equivalent if we cannot distinguish between them *externally*. That is to say, if we perform identical *external* measurements on both N_1 and N_2, we will always obtain identical results. In particular, if we connect the black boxes N_1 and N_2 separately to the *same* arbitrary network N, as shown in Fig. 7-1*a* and *b*, the values of the voltage and current solutions of the elements inside N must be identical in both cases. Let us now render the concept precise by the following definition.

DEFINITION OF EQUIVALENT NONLINEAR NETWORKS

Two n-terminal resistive black boxes N_1 and N_2 are equivalent if they can be interchangeably connected to the same *arbitrary* n-terminal network N without affecting the values of the voltages and currents *inside* N.†

We cannot overemphasize the fact that only the voltage and current solutions inside the testing network N need be identical. Nothing is said about the values of voltages and currents inside the black boxes N_1 and N_2 because these black boxes may contain completely different networks, and no comparison is meaningful nor desirable. In fact, the key idea here is that both N_1 and N_2 are really black inside, and all we can compare are the voltages and currents associated with the external network. A simple example will help to clarify this point. Consider the two networks N_1 and N_2 shown in the left side of Fig. 7-1*c* and *d*. If we connect an external network N consisting of a 3-Ω resistor across N_1 and N_2, we obtain the same solution ($i = 2$ amp and $v = 6$ volts) in each case. In fact, the solutions inside N will be the same regardless of what N is. Hence, we say N_1 and N_2 are equivalent. A common mistake committed by students is to conclude a priori that since N_1 is equivalent to N_2, the current and voltage associated with R_1 and R_2 must be equal. This is wrong because $i_1 = 2$ but $i_2 = 6$. Similarly, $v_1 = 2$ but $v_2 = 6$.

---

† This definition can be easily extended to include dynamic black boxes.

Exercise 1: Show by a counter-example that N_1 and N_2 are equivalent if, and only if, the definition is satisfied by *any* external network N, not just one particular network.

Exercise 2: Show that the definition of equivalent networks implies two properties: (*a*) Symmetric property: If N_1 is equivalent to N_2, then N_2 is equivalent to N_1, and hence the two networks may be interchanged, independent of the external load. (*b*) Transistive property: If N_1 is equivalent to N_2 and N_2 is equivalent to N_3, then N_1 is equivalent to N_3.

7-3 EQUIVALENCE BASED ON IDENTICAL DP PLOTS

We shall now derive a number of useful properties and theorems based on the principle that a two-terminal resistive black box N_1 is equivalent to a two-terminal resistive black box N_2 if, and only if, they have identical DP plots at all times.

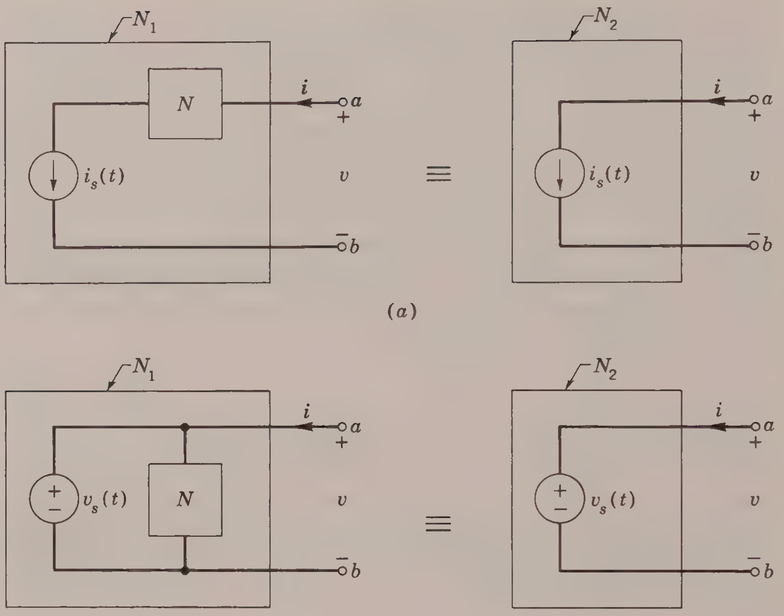
SERIES (PARALLEL) COMMUTATIVE PROPERTY

The DP plot across any resistive network containing two or more black boxes in series (parallel) is independent of the order in which the black boxes are connected, provided the relative orientation of each black box with respect to the driving-point terminals is preserved.

This property follows immediately from the series (parallel) combination techniques in Chap. 6.

REDUNDANT SERIES (SHUNT) ELEMENT PROPERTY

Any two-terminal black box N_1 which is connected in series (parallel) with a current (voltage) source is redundant, and may be replaced by a short (open) circuit without affecting the solution of any external network connected across the resulting black box N_2.



(*a*)

(*b*)

Fig. 7-2. The two black boxes N_1 and N_2 are equivalent because of the redundant series (shunt) element property.

The proof of this property follows from an inspection of the black boxes shown in Fig. 7-2*a* and *b*. The DP plots across N_1 and N_2 of Fig. 7-2*a* are identical because $i = i_s(t)$ for all t. Similarly, the DP plots across N_1 and N_2 of Fig. 7-2*b* are identical because $v = v_s(t)$ for all t.

As an application of this property, we can state the following corollary:

EQUIVALENT CURRENT (VOLTAGE) SOURCE PROPERTY

Any black box containing two or more identical current (voltage) sources in series (parallel) is equivalent to one current (voltage) source having the same terminal current (voltage).

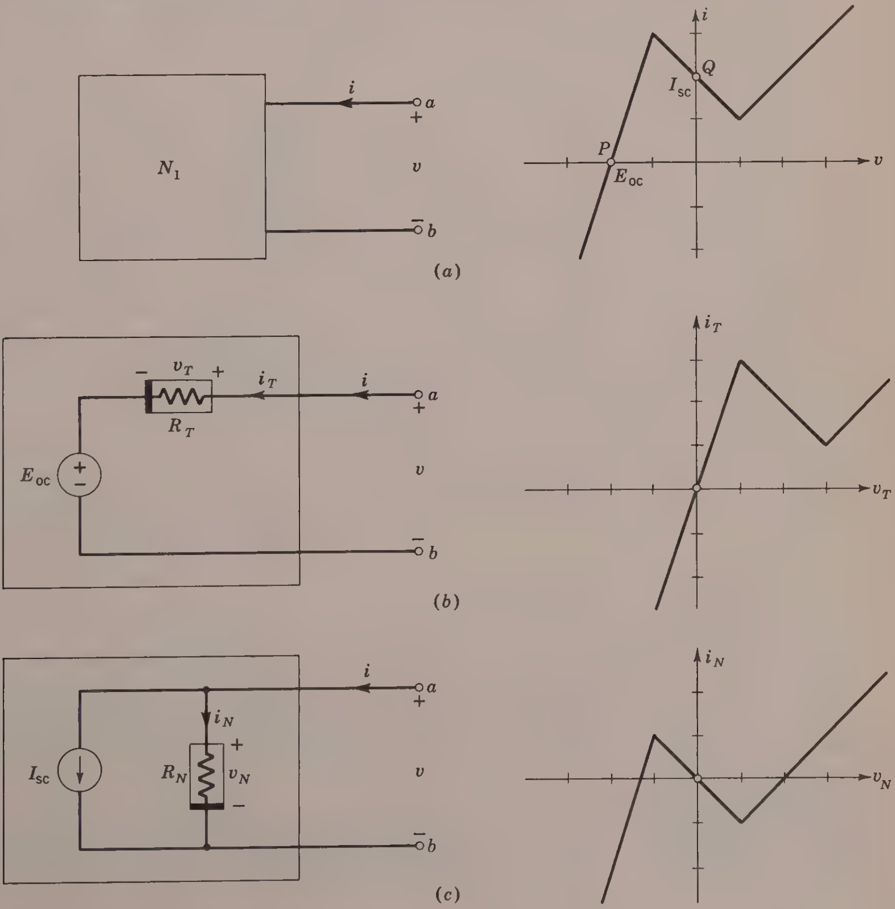
NONLINEAR THÉVENIN'S (NORTON'S) THEOREM

Any two-terminal dc-resistive black box is equivalent to an appropriate voltage (current) source in series (parallel) with an appropriate nonlinear resistor. The terminal voltage (current) of the voltage (current) source is equal to the voltage (current) intercept of the DP plot of the black box. The DP plot of the nonlinear resistor is obtained by translating the DP plot of the black box along its voltage (current) axis until it passes through the origin.

To illustrate this theorem, consider the black box N_1 and its DP plot as shown in Fig. 7-3*a*. The equivalent black box N_1 corresponding to Thévenin's and Norton's theorems is shown in Fig. 7-3*b* and *c*, respectively. Notice that the v_T-i_T curve in Fig. 7-3*b* is obtained by translating the origin of the DP plot in Fig. 7-3*a* to point P. Similarly, the v_N-i_N curve in Fig. 7-3*c* is obtained by translating the origin of the DP plot in Fig. 7-3*a* to point Q.

The proof of this theorem is obvious from an inspection of the respective DP plots in Fig. 7-3. Notice that the voltage E_{oc} at point P is equal to the terminal voltage v when $i = 0$. Since this voltage can be obtained by measuring the voltage across terminals *a*-*b* when these terminals are left open-circuited, it is usually called the *open-circuit voltage.* In contrast with this, the current I_{sc} is usually called the *short-circuit current* because it represents the current flowing up a short circuit connected across terminals *a*-*b*.

Fig. 7-3. An illustration of the application of the nonlinear Thévenin theorem and the nonlinear Norton theorem.



Exercise 1: Show that the equivalent current (voltage) source property follows from the redundant series (shunt) element property.

Exercise 2: Show that if the black box N_1 in Fig. 7-3a contains only linear elements and sources, then the resistors R_T and R_N in Fig. 7-3b and c are identical with a resistance value equal to $-E_{oc}/I_{sc}$.

7-4 EQUIVALENCE BASED ON IDENTICAL OPERATING POINTS

Consider the two black boxes N and N' in Fig. 7-4a. Suppose we pick N' to be the load and draw the hypothetical load line as shown in Fig. 7-4c. Let the voltage and current at the operating point be given by $v = E_Q$ and $i = I_Q$. Observe next that it is pos-

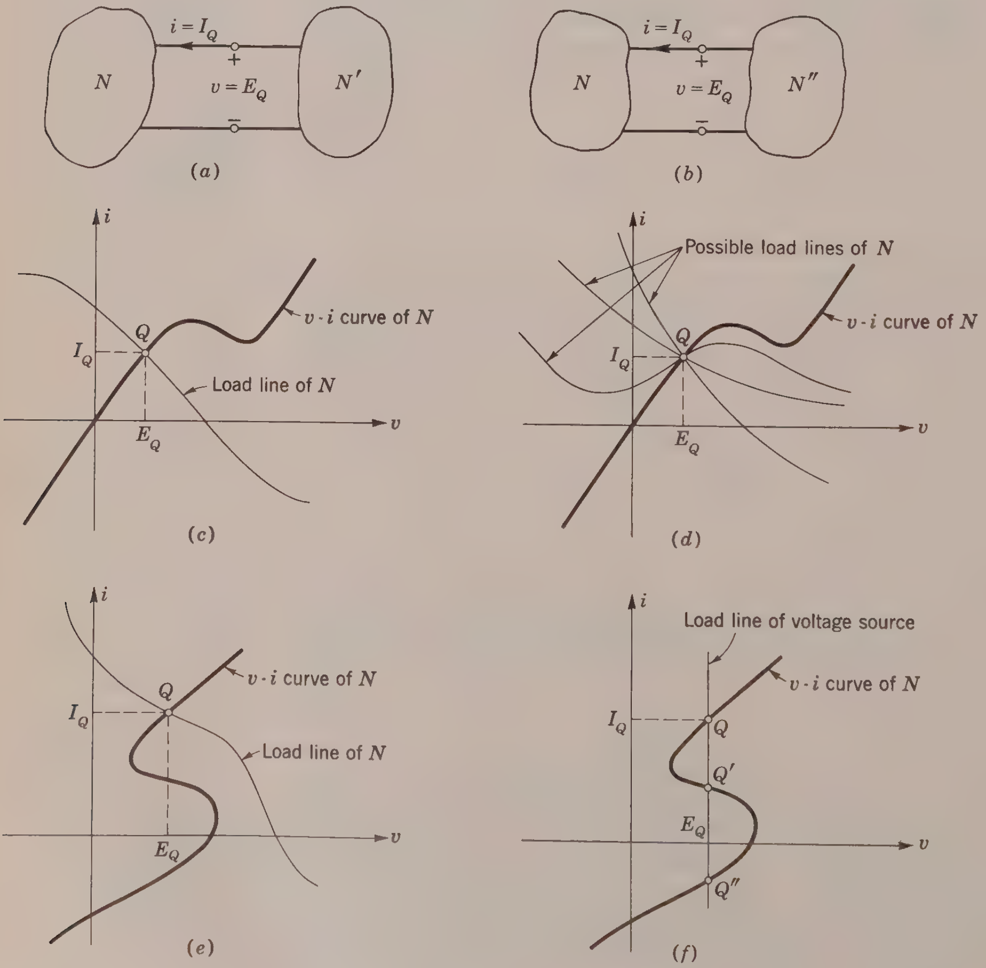


Fig. 7-4. An illustration of two black boxes N' and N'' having distinct DP plots but being equivalent *with respect to N* because both lead to an identical operating point.

sible to find many distinct load lines which lead to the same operating point. This means that it is possible to replace the load N' in Fig. 7-4*a* by a different load N'' in Fig. 7-4*b* without affecting in any way the voltage and current solutions of the elements inside the black box N. In fact, any of the load lines shown in Fig. 7-4*d* would qualify.[1] It is important to observe that the black boxes N' and N'' are *not* equivalent to each other because they have different DP plots. However, they can be said to be equivalent with respect to the black box N. This is a more restricted form of equivalence, but it is useful in simplifying analysis, as shown by the following substitution theorems:

VOLTAGE-SOURCE SUBSTITUTION THEOREM

If at any time t the black box N has a voltage-controlled DP plot, and if the terminal voltage is given by $v = v_s(t)$, then we can substitute the load N' by a voltage source with terminal voltage $v_s(t)$ without affecting the solution inside the black box N.

This theorem follows from the fact that a unique operating point results from the intersection between any voltage-controlled DP plot and a vertical load line. By assuming the same terminal voltage $v_s(t)$ for the voltage source, the same operating point will be obtained at all times. This proves the theorem. To show that the DP plot must be voltage-controlled in order for this theorem to be valid in general, suppose the v-i curve of N is not voltage-controlled, and suppose the load-line construction leads to the operating point Q as shown in Fig. 7-4*e*. If we replace the load N' by a voltage source with terminal voltage $v = E_Q$, the resulting load line will intersect the v-i curve of N not only at Q but also at Q' and Q'' as shown in Fig. 7-4*f*. In other words, if the DP plot of N is not voltage-controlled, extraneous solutions may result. By similar reasoning, we can now state the dual theorem:

CURRENT-SOURCE SUBSTITUTION THEOREM

If at any time t the black box N has a current-controlled DP plot, and if the terminal current is given by $i = i_s(t)$, then we can substitute the load N' by a current source with terminal current $i_s(t)$ without affecting the solution inside the black box N.

Since a short circuit is equivalent to a voltage source with zero terminal voltage, and an open circuit is equivalent to a current source with zero terminal current, we can now state the following theorems:

---

[1] If the original circuit has n operating points, the new circuit must also have n operating points.

SHORT-CIRCUIT SUBSTITUTION THEOREM

If the DP plot across a pair of terminals formed by a soldering-iron entry is voltage-controlled, and if the open-circuit voltage across these terminals is zero, then we can short-circuit the terminals without affecting the solution of the original network.

OPEN-CIRCUIT SUBSTITUTION THEOREM

If the DP plot across a pair of terminals formed by a pliers-type entry through a wire is current-controlled, and if the short-circuit current is zero, then we can open-circuit the wire without affecting the solution of the original network.

Exercise 1: Prove the current-source substitution theorem and show why the condition that the DP plot be current-controlled is necessary in general.

Exercise 2: Prove the short-circuit and the open-circuit substitution theorems and show why the condition on the DP plot is necessary in general.

Exercise 3: Show that the condition on the DP plot in the four substitution theorems can be replaced by a weaker condition, namely, the solution after the substitution remains unique.

7-5 THE SHIFTING THEOREMS

We are now ready to derive two important results which will allow us to break the series-parallel barrier. These are the v-shift theorem and the i-shift theorem. Since these theorems are crucial to many subsequent discussions in this book, the reader is strongly urged to develop a full understanding of these results.

THE v-SHIFT THEOREM

A voltage source connected across any pair of nodes a-b (such as Fig. 7-5a) can be shifted in series with *all* branches connected to either node a (Fig. 7-5b), thereby eliminating node a, or to node b (Fig. 7-5c), thereby eliminating node b. The polarity of the shifted sources at the common node must be identical with the polarity of the original source at the corresponding node.[1]

The v-shift theorem allows us to eliminate one node from the original network without affecting its solution. Very often, this

[1] Although only three branches are shown connected with each node, the theorem is valid for any number of branches.

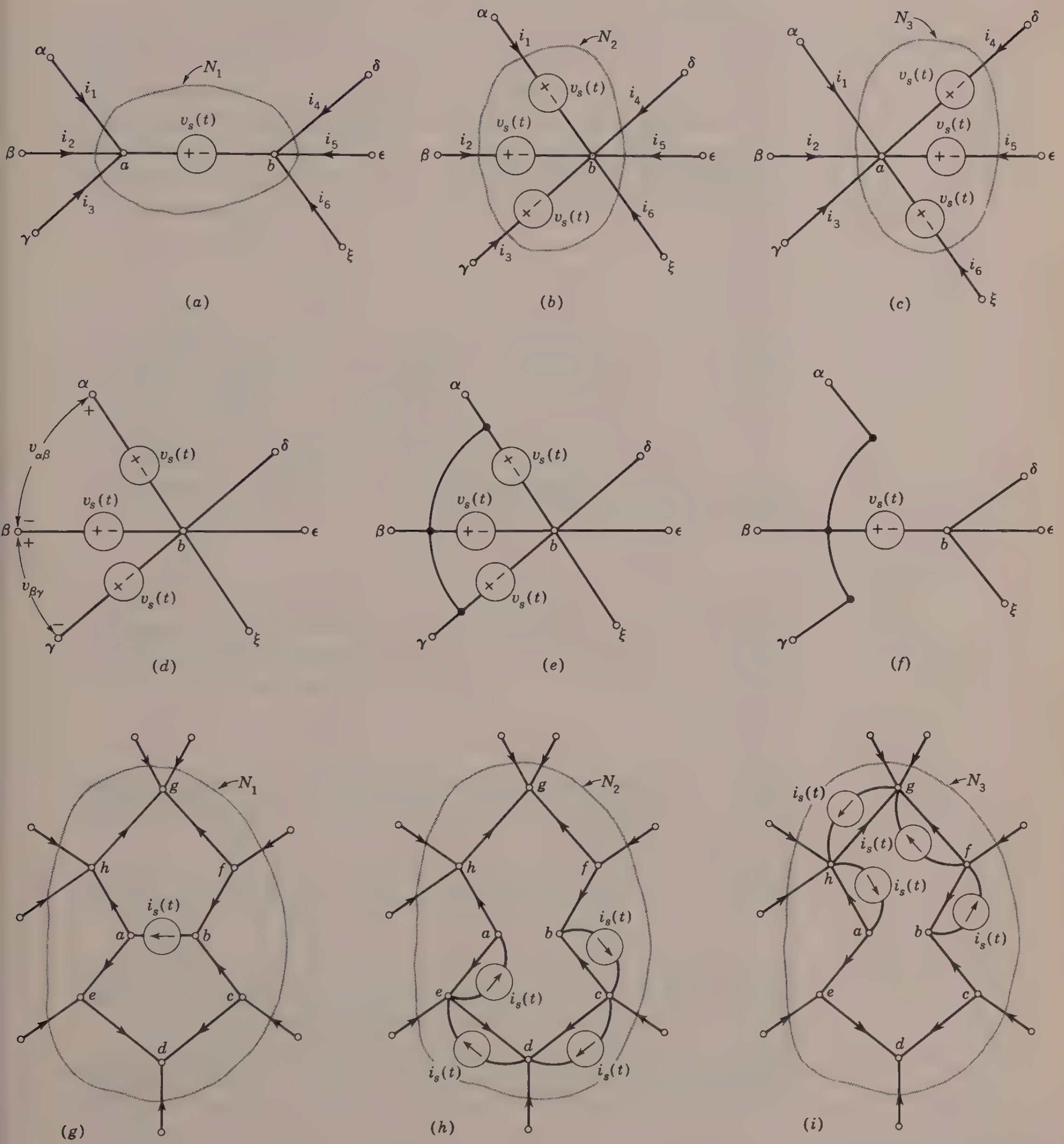


Fig. 7-5. An illustration of the removal of a node by the v-shift theorem and the removal of a branch by the i-shift theorem.

will transform a non-series-parallel network into one which is series-parallel. Before we apply this theorem to specific examples, let us prove it. It suffices to prove that the black box N_2 in Fig. 7-5b is equivalent to the original black box N_1, and hence they can be interchanged. For convenience, the black box N_2 is redrawn in

Fig. 7-5*d*. Observe that $v_{\alpha\beta} = v_{\beta\gamma} = 0$ and therefore we may connect a short circuit across terminals α, β, and γ (Fig. 7-5*e*) without affecting the original solution. Since the voltage sources in Fig. 7-5*e* are now in parallel, they can be replaced by one equivalent voltage source as shown in Fig. 7-5*f*. Observe that this is identical with the original network in Fig. 7-5*a*, and hence the theorem is proved.

THE *i*-SHIFT THEOREM

A current source connected across any pair of nodes *a-b* (such as Fig. 7-5*g*) can be shifted in parallel with any set of branches which form a closed loop with the original current source, such as loop *bcdea* in Fig. 7-5*h* or loop *bfgha* in Fig. 7-5*i*. The reference direction of the shifted sources is clockwise (counterclockwise) around the loop if the direction of the original source is in a counterclockwise (clockwise) direction.

The *i*-shift theorem allows us to remove a current source across a pair of nodes. This will often transform a non-series-parallel network into one which is series-parallel. Since the *i*-shift theorem is simply the dual of the *v*-shift theorem, its proof will be left as an exercise.

Exercise 1: Another way to prove the *v*-shift theorem is to show that the equations of motion of the shifted network remain unchanged. Verify this property for both N_2 and N_3 in Fig. 7-5*b* and *c*.

Exercise 2: Prove the *i*-shift theorem by two methods: (*a*) Construct the dual of the proof given for the *v*-shift theorem. (*b*) Show that the equations of motion remain unchanged.

Exercise 3: The *v*-shift theorem can be generalized by shifting the voltage source across any pair of nodes *a-b* to all branches which form a cut set with the voltage source. State this generalized theorem precisely and prove it.

Exercise 4: Are the shifting theorems applicable to controlled sources? If so, what precaution must be taken?

7-6 GRAPHICAL ANALYSIS OF NON-SERIES-PARALLEL NETWORKS

We shall now demonstrate by means of examples the application of equivalent network theorems to the analysis of a large class of non-series-parallel networks.

EXAMPLES

1. We are now in a position to solve for the operating point of the network shown in Fig. 6-12, which we reproduce in Fig. 7-6*a* for convenience. It was shown earlier in Sec. 6-2-2 that the double load-line method is applicable provided the DP plot across the black boxes N_1 and N_2 can be found.

   Consider first the black box N_1. The presence of the voltage source v_{11} suggests that we apply the v-shift theorem to transform N_1 into the equivalent series-parallel network shown in Fig. 7-6*b*. The v_3-i_3 DP plot of this network can now be easily obtained by the methods described in Chap. 6. Consider next the black box N_2. This time let us apply the i-shift theorem to obtain the equivalent series-parallel network shown in the center of Fig. 7-6*c*. The DP plot can be obtained by a similar procedure. We can save some work, however, by observing that the voltage source v_{12} is connected across an imaginary two-terminal black box made up of resistors R_6 and R_8 and the current source I. Hence, applying the redundant shunt element property (Sec. 7-3), we can remove resistors R_6 and R_8 and the current source I without affecting the v_4-i_4 DP plot of N_2. This is shown in the right of Fig. 7-6*c*.

   Assuming that the above procedures have been carried out and the DP plots across N_1 and N_2 have been found, we can proceed to determine the operating-point voltages v_3 and v_4 and currents i_3 and i_4 by the double load-line method. Let us suppose that this, too, has been done, and consequently we found $v_3 = E_3$, $v_4 = E_4$, $i_3 = I_3$, and $i_4 = I_4$. We can now "enter" each black box and determine the solution of each element as follows. If the DP plot across N_1 is voltage-controlled, we can apply the voltage-source substitution theorem (Sec. 7-4) and replace the network to the right of N_1 by a voltage source with a terminal voltage equal to E_3 as shown in the left of Fig. 7-6*d*. If the DP plot across N_1 is current-controlled, we can apply the current-source substitution theorem and replace the network to the right of N_1 by a current source with a terminal current equal to I_3 as shown in the right of Fig. 7-6*d*. The same principle holds for the networks shown in Fig. 7-6*e*.[1] Replacing the black boxes N_1 and N_2 by the equivalent series-parallel networks in Fig. 7-6*b* and *c*, respectively, we can easily determine the operating-point voltage and current associated with each element inside the black boxes N_1 and N_2.

---

[1] If the DP plot of either N_1 or N_2 is multivalued, the principle still holds. However, we must watch for the possibility of extraneous solutions so that they can be excluded in the final solution.

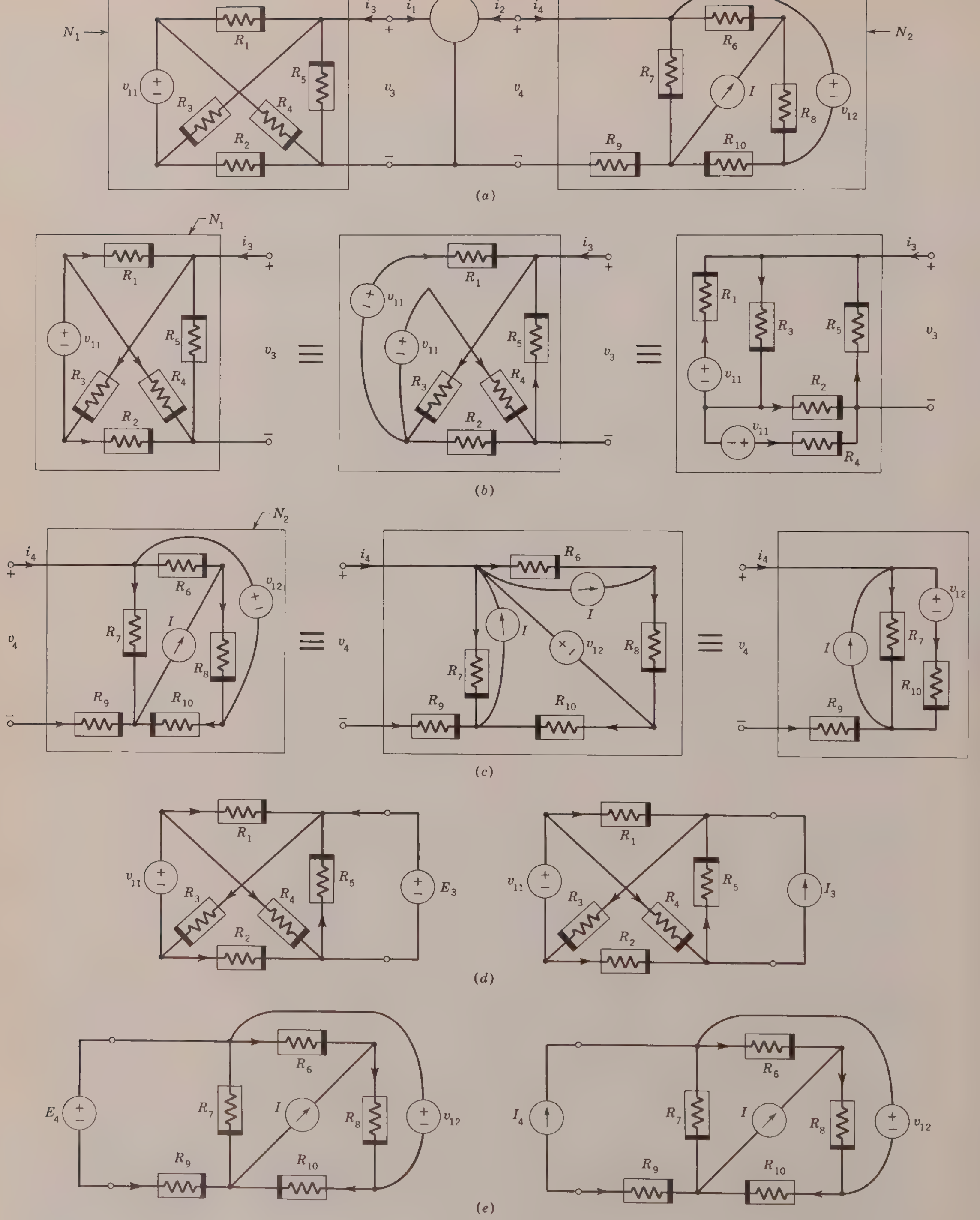


Fig. 7-6. An illustration of the application of the shifting theorems and the substitution theorems for analyzing a complicated non-series-parallel network.

2. To show that the shifting theorems are applicable not only to networks with built-in voltage sources or current sources, such as the network in Fig. 7-6*a*, consider the nonlinear *lattice network* shown in Fig. 7-7*a*.[1] The basic idea is to recast this network into the form of basic configuration 2 of Sec. 6-2-2 as shown in Fig. 7-7*b*. We can apply the double load-line method to find the v_o-vs.-v_i TC plot of this network if we can represent the black box N by two sets of characteristic curves in the same manner in which a three-terminal resistor is represented.

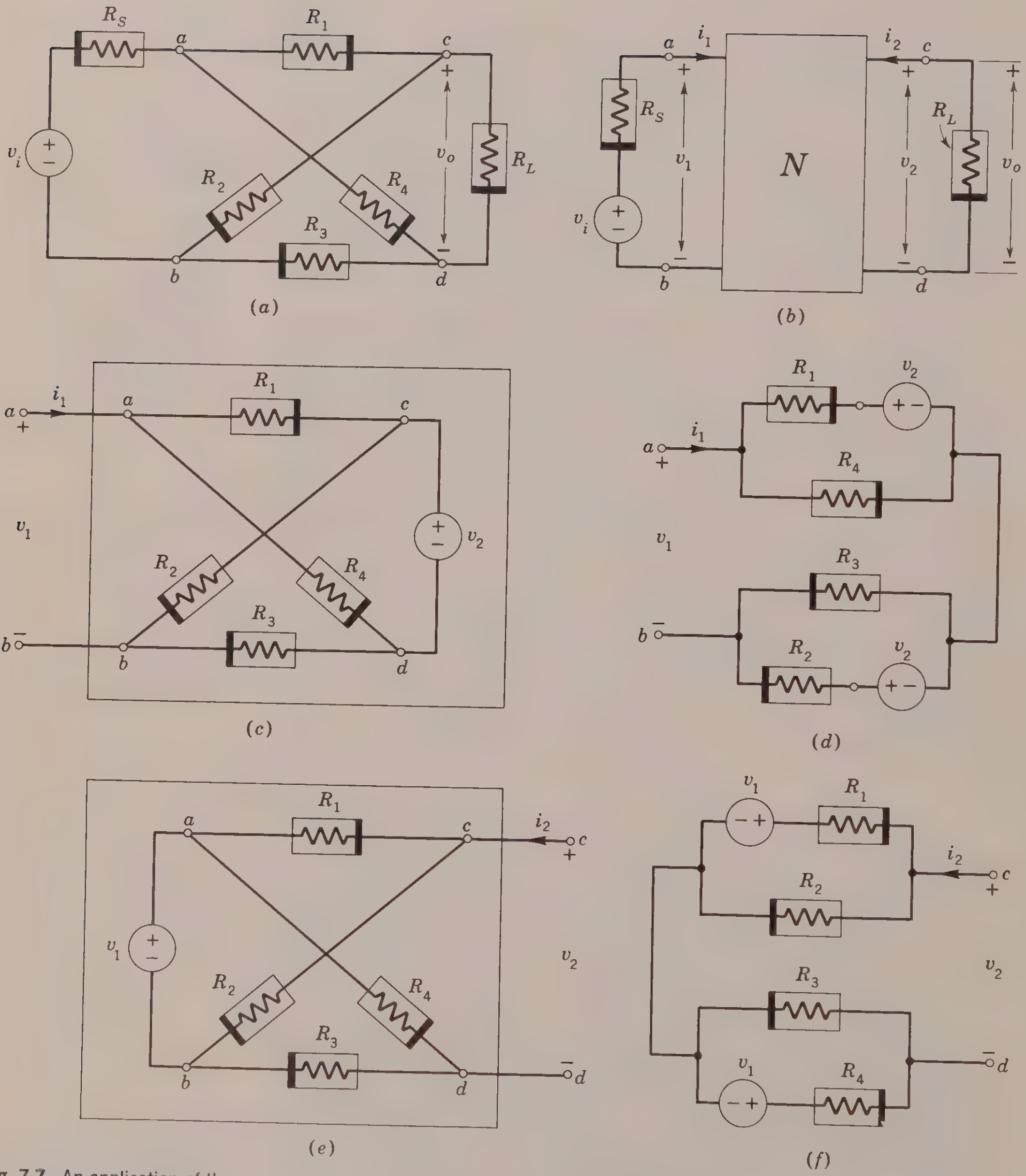
   To obtain the set of input characteristics in the conductance representation, we extract the black box N and connect a voltage source v_2 across terminals *c-d* as shown in Fig. 7-7*c*. We then apply the v-shift theorem to obtain the equivalent series-parallel network shown in Fig. 7-7*d*. For each value of the parameter v_2, we obtain one DP plot. Hence, by repeating this procedure for several values of v_2, we obtain the desired set of input-characteristic curves.

   To obtain the set of output characteristics in the conductance representation, we extract the black box N and connect a voltage source v_1 across terminals *a-b* as shown in Fig. 7-7*e*. An application of the v-shift theorem results in the equivalent series-parallel network shown in Fig. 7-7*f*. The set of output-characteristic curves is obtained by finding the DP plot across terminals *c-d* corresponding to different values of the parameter v_1.

   As a simple application of the above general procedure, consider the full-wave rectifier circuit shown in Fig. 7-8*a*. The most important information desired is the TC plot between the output voltage v_o and the input voltage v_i. We would also like to know the effect of the load resistance R_L on the resulting TC plot. Since the resistor R_s in Fig. 7-7*a* is absent in this circuit, the TC plot can be simply obtained by finding only the v_2-i_2 output-characteristic curves of the equivalent circuit shown in Fig. 7-8*b* as a function of the input voltage parameter v_i. Assuming a typical diode v_d-i_d curve as shown in Fig. 7-8*c*, the v_2-i_2 curves are easily obtained graphically as shown in Fig. 7-8*d*. Notice that for this particular circuit, each curve is associated with both $v_i = E$ and $v_i = -E$. We can now draw the load line corresponding to R_L and obtain the TC plot by transferring the points of intersection into the v_i-v_o plane. The TC plots corresponding to three different load resistances are shown in Fig. 7-8*e*. Observe that as the value of R_L decreases, the output voltage v_o also decreases.

---

[1] This network configuration is chosen not only because there are many practical circuits having the same topology, but also because the same idea can be applied to many other network configurations.

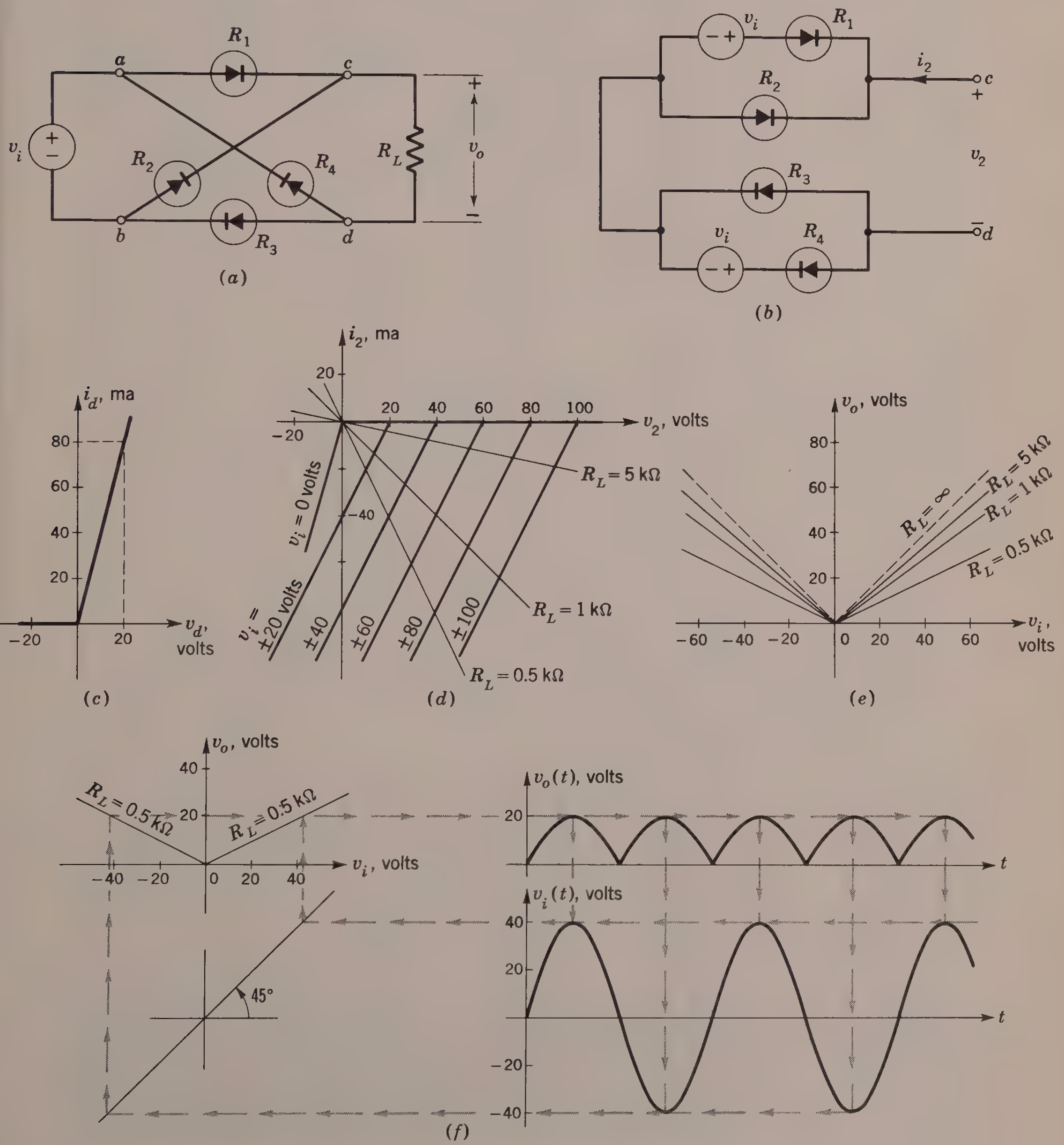


(a) (b) (c) (d) (e) (f)

Fig. 7-7. An application of the v-shift theorem for determining the TC plot of a nonlinear lattice network.

The output voltage waveform $v_o(t)$ corresponding to a sinusoidal input voltage $v_i(t)$ is obtained by the graphical construction shown in Fig. 7-8*f*. Notice that both positive and negative waves are now present in the rectified output voltage, hence the name full-wave rectifier. Ideally, we would like to have the peak output voltage equal the peak input voltage. Accordingly, the ideal TC plot should be described by $v_o = |v_i|$. It is clear that the nonlinearity of the diode v_d-i_d curve and the finite load resistance are responsible for the departure from this ideal TC plot.

Fig. 7-8. A detailed graphical analysis of the full-wave rectifier.



3. The shifting theorems are particularly useful in the determination of the characteristic curves of a three-terminal black box containing both two-terminal and three-terminal resistors.

   Consider first the configuration shown earlier in Fig. 6-43*c*. This is reproduced in Fig. 7-9*a* for convenience. To determine the input-characteristic curve corresponding to each value $i_b = k$ of the current source connected across terminals *b-c*, we apply the *i*-shift theorem to obtain the equivalent network shown in Fig. 7-9*b*. Notice that this circuit can be considered as a series connection between the two black boxes N_1 and N_2, and hence the DP plot corresponding to each value $i_b = k$ is easily obtained. To determine the output-characteristic curve corresponding to each value $i_a = k$ of the current source connected across terminals *a-c*, we apply the *i*-shift theorem and obtain the equivalent circuit shown in Fig. 7-9*c*. The DP plot corresponding to each value $i_a = k$ is obtained by a similar procedure.

   Consider next the configuration shown earlier in Fig. 6-43*d*. This is reproduced in Fig. 7-9*d* for convenience. To obtain the input-characteristic curve corresponding to each value $v_{bc} = k$ of the voltage source connected across terminals *b-c*, we apply the *v*-shift theorem to obtain the equivalent network shown in Fig. 7-9*e*. Notice that this circuit can be considered as a parallel connection between the two black boxes N_1 and N_2, and hence the DP plot corresponding to each value $v_{bc} = k$ is easily obtained. To determine the output-characteristic curve corresponding to each value $v_{ac} = k$ of the voltage source connected across terminals *a-c*, we apply the *v*-shift theorem and obtain the equivalent circuit shown in Fig. 7-9*f*. The DP plot corresponding to each value $v_{ac} = k$ is obtained by a similar procedure.

   As a final illustration of the utility of the shifting theorems, consider the famous Darlington transistor connection shown in Fig. 7-9*g*. This circuit is widely used as a three-terminal device because it offers a number of advantages over a single transistor. To determine the characteristic curves of this three-terminal black box, it is convenient to apply a voltage source across terminals *b-c* and a current source across terminals *a-c*. Corresponding to each value $v_{bc} = k$ of the voltage source, the input-characteristic curve is determined by applying the *v*-shift theorem to obtain the equivalent network shown in Fig. 7-9*h*. Notice that the black box *N* can be interpreted as the resistor *R* in the circuit configuration of Fig. 7-9*a*. Hence, the technique discussed

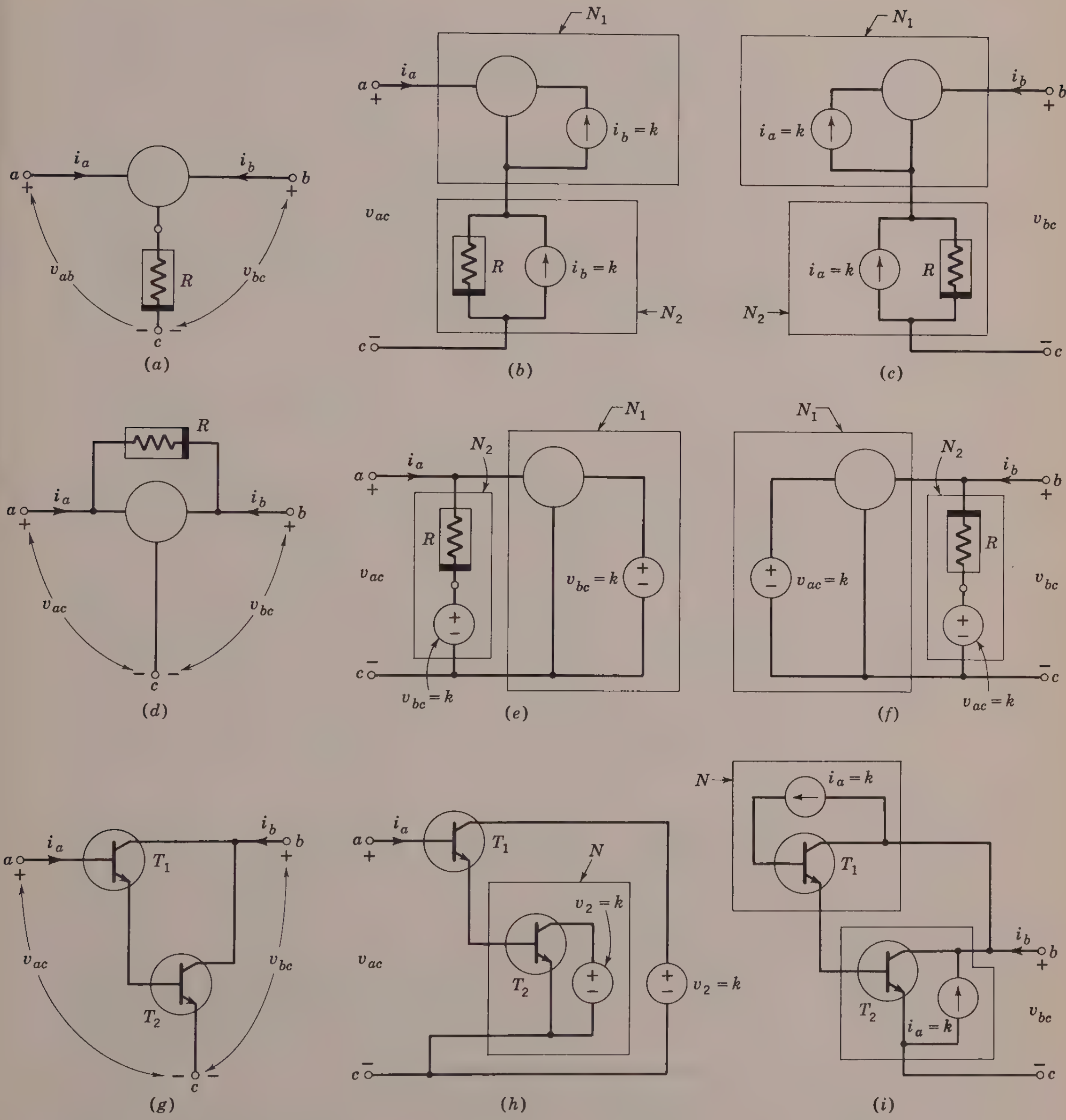


Fig. 7-9. Application of the shifting theorems to obtain the characteristic curves of three-terminal resistive black boxes.

above for this configuration is applicable here. To determine the output-characteristic curve corresponding to each value $i_a = k$, we apply the *i*-shift theorem to obtain the equivalent circuit shown in Fig. 7-9*i*. Observe that the black box *N* can be interpreted as the resistor R in the circuit configuration of Fig. 7-9*d*. Hence, the technique discussed above for this configuration is again applicable here.

The above procedure can obviously be used to obtain the characteristic curves of three-terminal black boxes containing more than two transistors connected in the Darlington configuration or other configurations. The basic principle is to identify appropriate black boxes, through the application of equivalent network theorems, so that the circuit reduces to a combination of the basic configurations discussed so far.[1]

Exercise 1: Develop a general method for determining the TC plot of the nonlinear lattice network shown in Fig. 7-7*a* with the help of the *i*-shift theorem.

Exercise 2: Describe the step-by-step graphical procedure for finding the exact TC plot and DP plot of the full-wave rectifier circuit shown in Fig. 7-8*a*, but with a linear resistor R_s inserted in series with the voltage source. You may assume a piecewise-linear *v*-*i* curve for the diode in carrying out the discussion.

Exercise 3: Compare the methods described above for determining the characteristic curves of the three-terminal black boxes shown in Fig. 7-9*a* and *d* with those described earlier in Sec. 6-8-1. What are the advantages and disadvantages in each case?

Exercise 4: Discuss the step-by-step procedure for determining the characteristic curves of the circuit in Fig. 7-9*g* but with each transistor T_1 and T_2 replaced by the Darlington pair of Fig. 7-9*g*.

7-7 PRINCIPLES AND APPLICATIONS OF SYMMETRY

Many difficult problems in physical science have been solved by rather simple methods. One such method is to take advantage of the symmetry exhibited by some systems. Applying the concept of symmetry to nonlinear networks, if indeed they are symmetrical, often leads to very simple solutions. Failure to take advantage of such symmetry can make the network much more difficult, and sometimes impractical, to analyze.

Roughly speaking, a network is symmetrical with respect to a certain physical transformation, such as a rotation, if the network remains invariant after the transformation. The word *invariant* is used in the strict sense to indicate that the transformed network remains exactly identical with the original network, *both geometrically and electrically*. We shall now state a precise definition concerning two rather common types of symmetry.

---

[1] A computer program can be written to implement the series of graphical procedures automatically.

DEFINITION OF FLIP-OVER AND ROTATIONAL SYMMETRIES

A network is said to be symmetrical with respect to an "axis of symmetry" α if it remains geometrically and electrically invariant upon a rotation of θ deg about α. If α lies on the plane of the network, then θ equals 180° and the network is said to exhibit a *flip-over symmetry.*[1] If α is a line perpendicular to the plane of the network, then the network is said to exhibit a θ-deg *rotational symmetry.*

To illustrate the various types of symmetry, Fig. 7-10*a* to *c* show three networks exhibiting flip-over symmetry corresponding to three different axes α_1, α_2, and α_3, respectively. These three types of flip-over symmetry can be further distinguished as vertical, horizontal, and skew symmetry. However, since any network exhibiting flip-over symmetry can be redrawn so that the axis α becomes vertical, it suffices to consider vertical flip-over symmetry. The networks shown in Fig. 7-10*d* to *f* exhibit rotational symmetry with $\theta = 180$, 120, and 60°, respectively. Observe that any network which exhibits θ-deg rotational symmetry also exhibits $n\theta$-deg rotational symmetry, where n is an integer. Hence, the network of Fig. 7-10*f* also exhibits 120° rotational symmetry.

It is important to realize that the black boxes shown in Fig. 7-10*a* to *f* must also exhibit electrical symmetry. For example, the network shown in Fig. 7-10*g* "seems" to exhibit vertical flip-over symmetry. If we actually flip over this network as shown in Fig. 7-10*h*, we find the network to be only geometrically symmetric. This is because the polarities of the two nonlinear resistors R are now reversed, and hence their v-i curves are different, unless the resistors are bilateral.[2] Indeed, if we redraw the network in Fig. 7-10*h* into the black-box configuration shown in Fig. 7-10*i*, it is clear that the DP plots of N' and N'' will not be identical unless the resistor R is bilateral.

7-7-1 SOME APPLICATIONS OF SYMMETRY TO THE OPERATING-POINT PROBLEM

We shall now derive a number of theorems which will often lead to shortcuts in the determination of the operating point of a symmetric network. In order for these theorems to be valid for all cases, it is necessary to stipulate the following standing assumption: *The network under consideration has a unique operating point.*[3] We also assume that *the short-circuit and the open-circuit substitution theorems are applicable whenever the voltage or the current is equal to zero.*

---

[1] This term is suggested by the observation that if the network is drawn on a transparent sheet of paper, and if we flip over the paper along the axis α, the network remains unchanged.

[2] As defined in Sec. 1-6-3, a resistor is said to be bilateral if, and only if, its v-i curve is symmetrical with respect to the origin.

[3] One sufficient condition is the uniqueness theorem in Sec. 5-6.

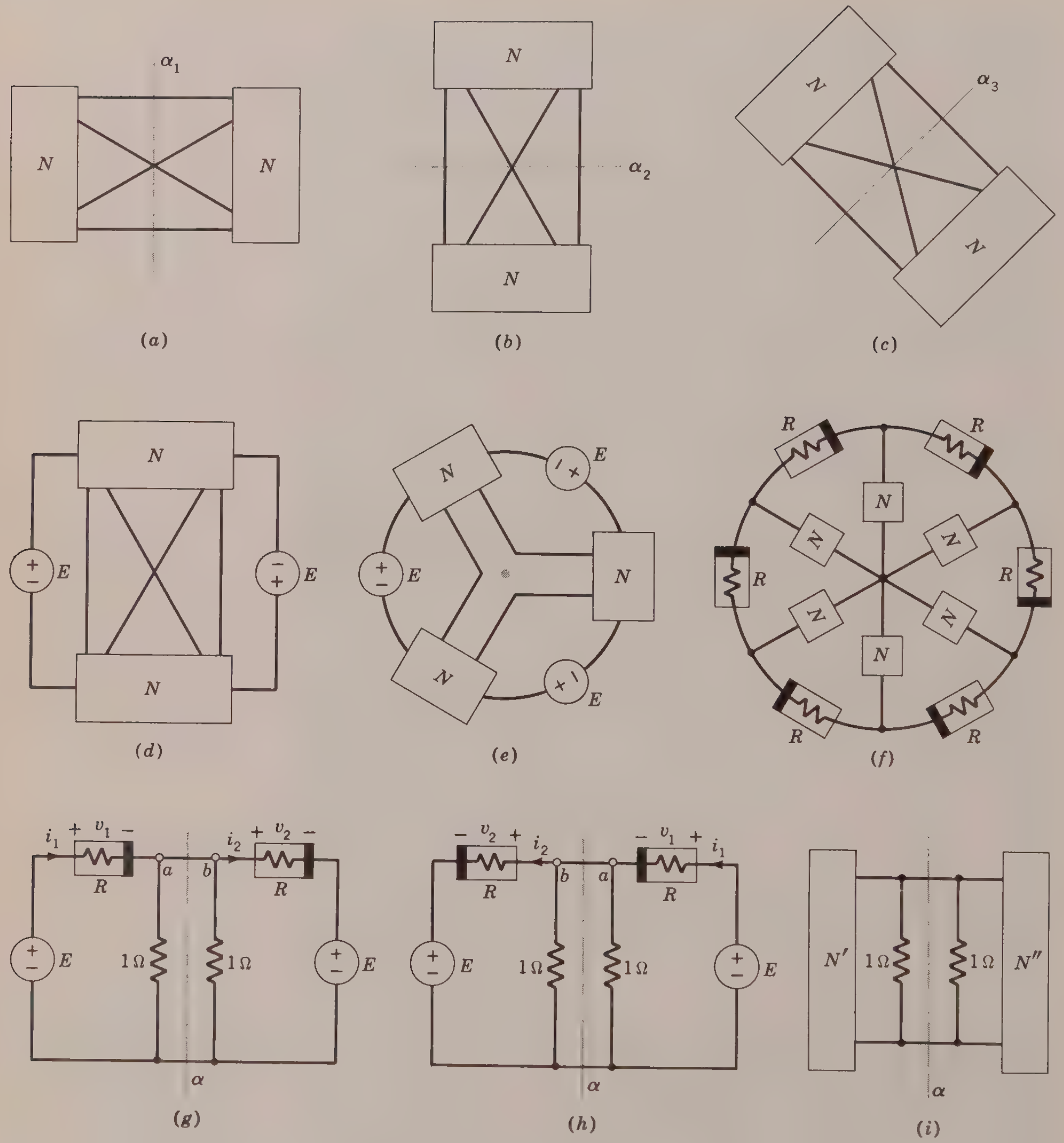


Fig. 7-10. An illustration of various types of symmetrical networks.

We shall henceforth refer to the above assumption as the *uniqueness hypothesis.* If this assumption is not satisfied, the following theorems may still be valid; however, one must be careful to watch in each step that no extraneous solutions are introduced.

FLIP-OVER SYMMETRIC-NETWORK THEOREM

Under the uniqueness hypothesis, the flip-over symmetric network shown in Fig. 7-11*a* is equivalent to the bisected network shown in Fig. 7-11*b*. Hence, all horizontal wires may be open-circuited, and all pairs of crossed wires may be short-circuited, without affecting the operating point of the network.

This theorem is intuitively obvious because if we flip over the network, we could not tell any difference. There is no preference, therefore, for the current in the horizontal wires to flow in one direction or the other. Similarly, there is no preference for the voltage across each side of the crossed wires to assume either polarity. The only way to satisfy this neutrality is for the currents and voltages to be zero. We could, of course, prove this theorem rigorously. To do this, it suffices to consider an arbitrary horizontal wire, such as *a-b* of Fig. 7-11*a*, and a pair of crossed wires, such as wires *c-d* and *e-f*. Let us now flip over the network and observe that point *a* becomes point *b*, and vice versa. Similarly, points *c* and *e* in the original network become points *f* and *d* in the flipped-over network, and vice versa. Since the network is assumed to have a unique solution, and since the two networks are identical, we have $i_{ab} = i_{ba}$ and $v_{ce} = v_{fd}$. But if we apply KCL and KVL to the original network in Fig. 7-11*a*, we obtain the relations $i_{ab} = -i_{ba}$ and $v_{ce} = -v_{fd}$. Since a number can equal its negative only if it is zero, we conclude that $i_{ab} = i_{ba} = 0$ and $v_{ce} = v_{fd} = 0$. Furthermore, in view of the uniqueness hypothesis, we can apply the open-

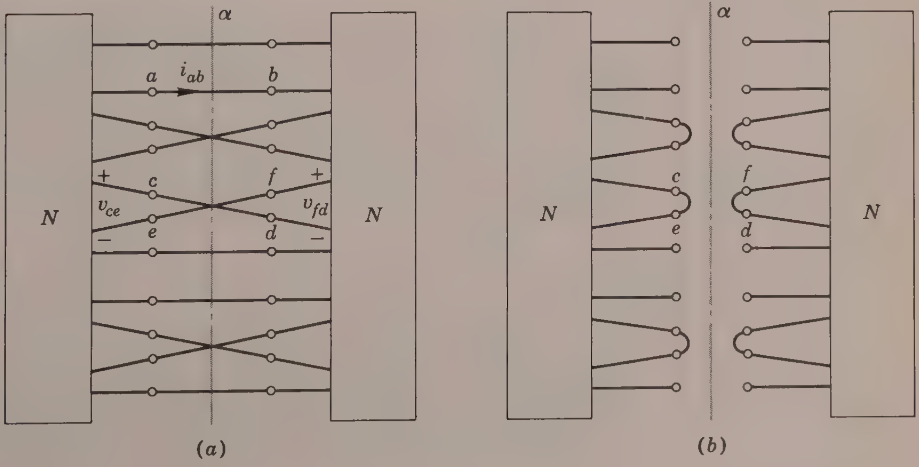


Fig. 7-11. Two equivalent networks exhibiting flip-over symmetry.

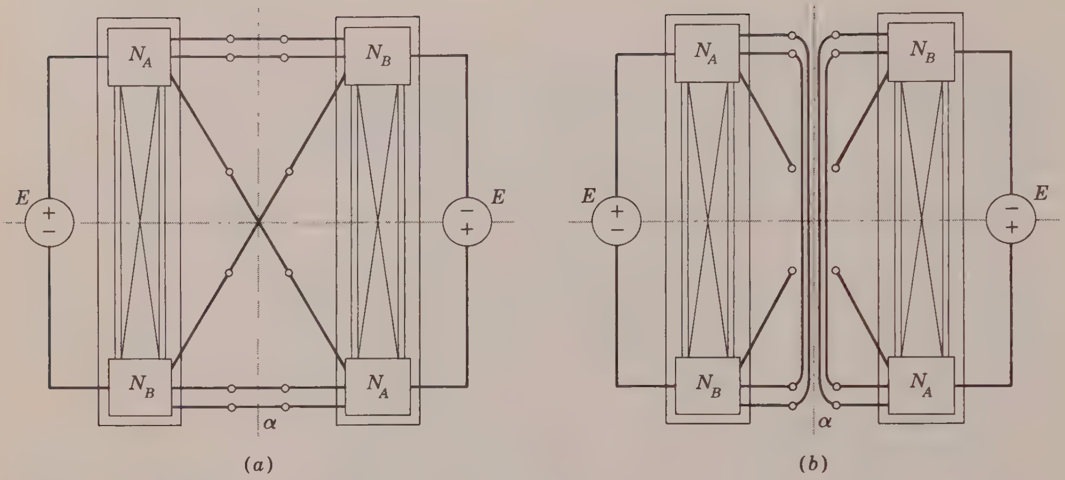


Fig. 7-12. Two equivalent networks exhibiting 180° rotational symmetry.

circuit substitution theorem and the short-circuit substitution theorem (Sec. 7-5) to obtain the equivalent network shown in Fig. 7-11*b*.

ROTATIONAL SYMMETRIC-NETWORK THEOREM

Under the uniqueness hypothesis, the 180° rotational symmetric network shown in Fig. 7-12*a* is equivalent to the bisected network shown in Fig. 7-12*b*. Hence, each pair of crossed wires may be open-circuited, and each pair of corresponding horizontal wires may be short-circuited, without affecting the operating point of the network.

This theorem follows from the observation that the network shown in Fig. 7-12*a* can be redrawn into one which exhibits vertical flip-over symmetry by "flipping" or "twisting" the right portion of the network along a horizontal axis through the center. In so doing, each pair of crossed wires becomes a pair of horizontal wires, and each pair of horizontal wires becomes a pair of crossed wires.

SYMMETRIC LATTICE NETWORK THEOREM

Under the uniqueness hypothesis, the voltages and currents of the two identical series-arm resistors R_A (crossed-arm resistors R_B) of the symmetric, nonlinear lattice network shown in Fig. 7-13*a* are equal to each other; that is, $v_1 = v_3$, $v_2 = v_4$, $i_1 = i_3$, and $i_2 = i_4$.

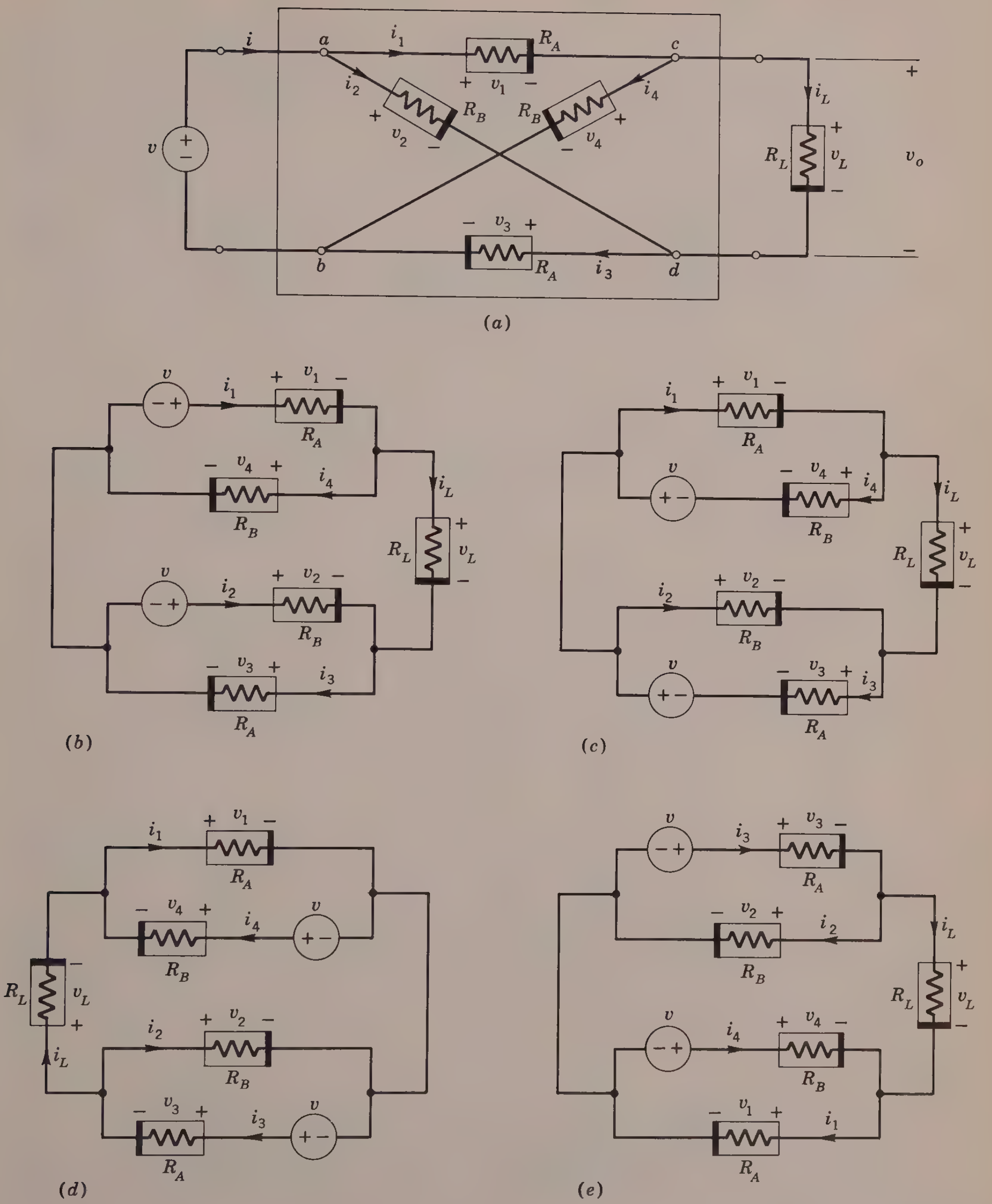


Fig. 7-13. The step-by-step proof of the symmetric lattice network theorem.

To prove this theorem, let us first apply the v-shift theorem through nodes a and b of the lattice network and obtain the equivalent networks shown in Fig. 7-13*b* and *c*, respectively. Let us next apply the series commutative property (Sec. 7-3) to move the resistor R_L from the network in Fig. 7-13*c* into the position shown in Fig. 7-13*d*. The same property is used to interchange the position of the voltage source v and the resistors R_A and R_B (Fig. 7-13*d*). Finally, let us rotate the network in Fig. 7-13*d* by 180° with respect to an axis perpendicular to the plane of the network as shown in Fig. 7-13*e*. Observe now that the networks shown in Fig. 7-13*b* and *e* are identical except for the voltage and current variables associated with the resistors. Moreover, in view of the uniqueness hypothesis, the network has a unique solution. Hence, we conclude that $v_1 = v_3$, $v_2 = v_4$, $i_1 = i_3$, and $i_2 = i_4$.

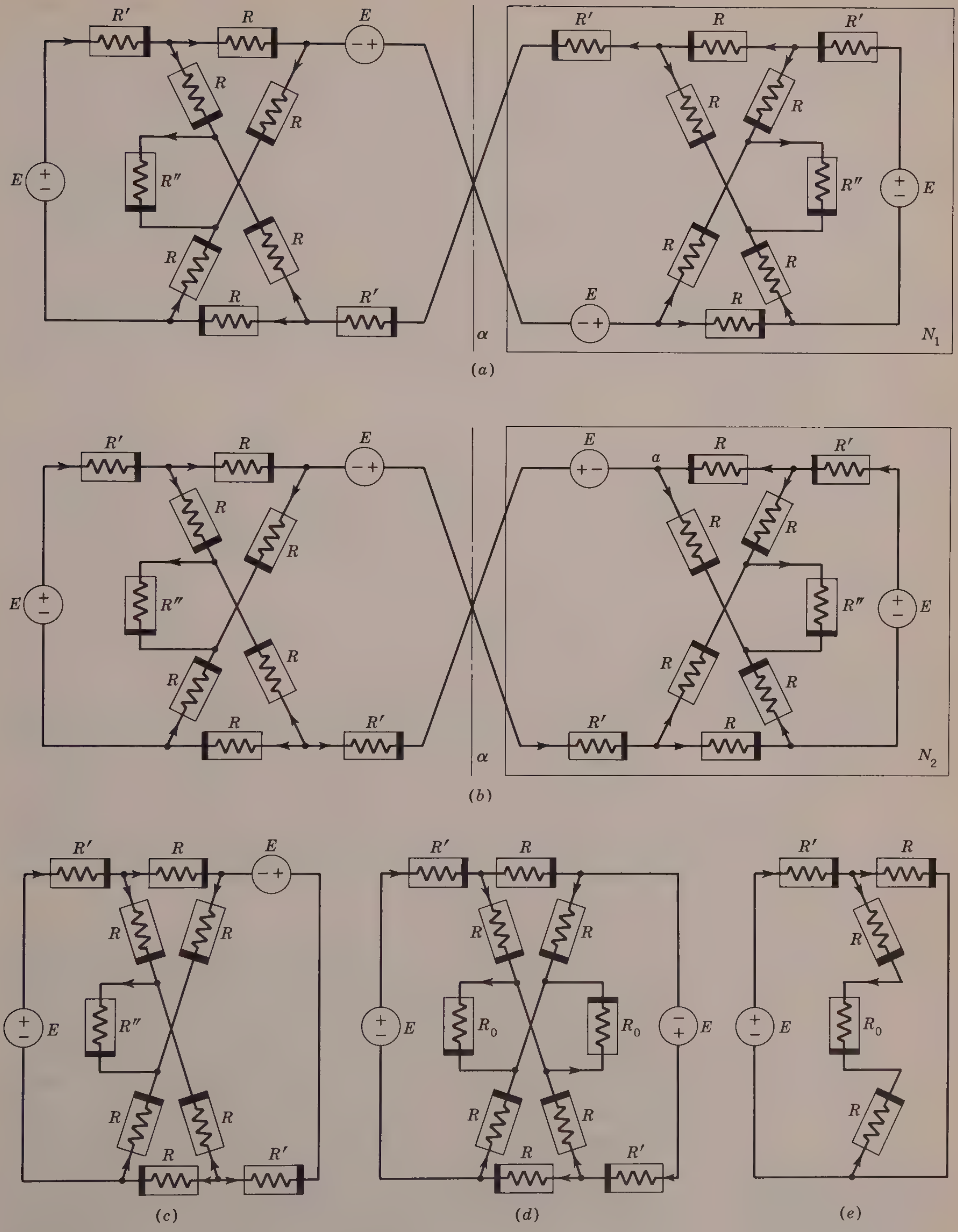
Whenever applicable, the above theorems greatly simplify the solution of many operating-point problems. As an example, consider the network shown in Fig. 7-14*a*, where all resistors are assumed to be characterized by monotonically increasing v-i curves. Moreover, we shall assume that the four resistors labeled R' are bilateral. An initial examination of this network does not reveal any form of symmetry. However, it is possible to transform this network so that it exhibits flip-over symmetry by interchanging the position of the resistor R' and the voltage source E as shown in Fig. 7-14*b*. Since the resistor R' is bilateral, we may transpose its terminals, thereby obtaining a flip-over symmetric network. Hence, applying the flip-over symmetric-network theorem, we can short the pair of crossed wires to obtain the equivalent network shown in Fig. 7-14*c*.[1] This network can be further simplified by splitting the resistor R'' into two identical resistors R_0 in parallel as shown in Fig. 7-14*d*. Notice that the terminals of the bilateral resistor R' have been transposed in this figure in order that the resulting network exhibit rotational symmetry. An application of the rotational symmetric-network theorem leads to the equivalent network shown in Fig. 7-14*e*. This network is now relatively simple to analyze.

7-7-2 SOME APPLICATIONS OF SYMMETRY TO THE DETERMINATION OF DP PLOTS AND TC PLOTS

There are many practical networks for which the DP plot or TC plot exhibits some form of symmetry. If we can tell, a priori, the type of symmetry, then only half of the work needs to be done. We shall now derive a number of useful properties for this purpose.

---

[1] Since the two bisected subnetworks are identical, only one part needs to be considered.

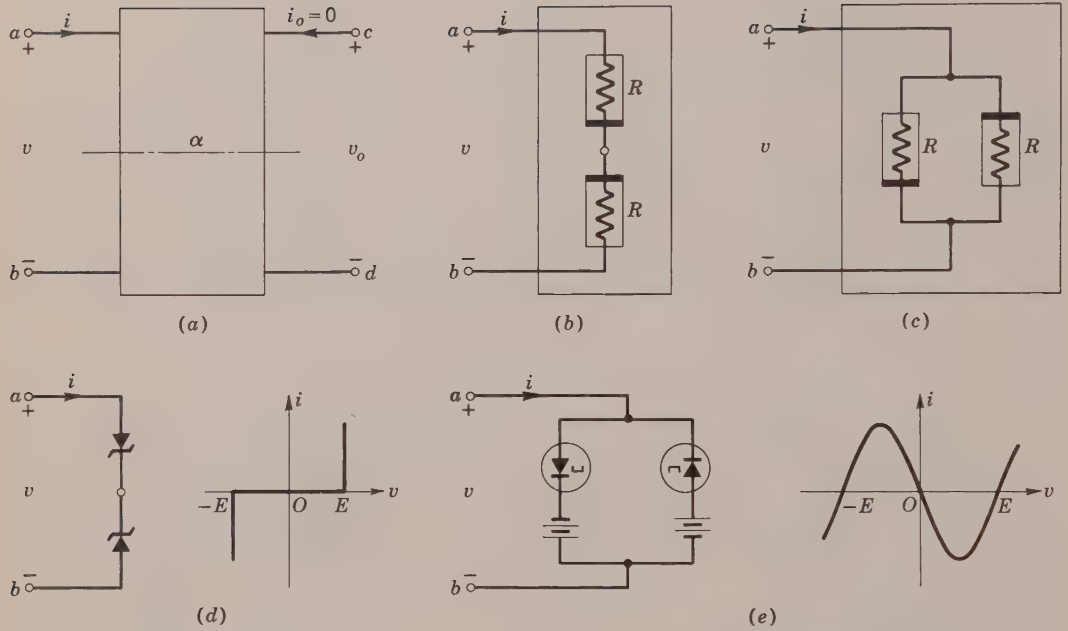


(a)

(b)

(c) (d) (e)

Fig. 7-14. An example illustrating the application of flip-over and rotational symmetries.



(a) (b) (c)

(d) (e)

Fig. 7-15. The DP plot and TC plot of any horizontally symmetric black box are symmetrical with respect to the origin.

SYMMETRIC DP PLOT–TC PLOT THEOREM

If a black box exhibits a flip-over symmetry with respect to the horizontal axis α through the center of the box (Fig. 7-15*a*), then the DP plot across any pair of symmetrically spaced driving-point terminals is symmetrical with respect to the origin. Similarly, the TC plot from any pair of symmetrically spaced output terminals to a pair of symmetrically spaced driving-point terminals is also symmetrical with respect to the origin.

The first part of the theorem follows immediately from the fact that the flip-over symmetry implies that the two-terminal black box is bilateral. To prove that the TC plot is also symmetrical, let us denote the TC plot of the black box in Fig. 7-15*a* by $v_{cd} = f(v_{ab})$. If we flip over the box, the TC plot remains unchanged because the two black boxes are indistinguishable. Hence, $v_{dc} = f(v_{ba})$. But KVL implies that $v_{cd} = -v_{dc}$ and $v_{ab} = -v_{ba}$. This means that $v_{cd} = -f(v_{ba}) = -f(-v_{ab})$, and hence the TC plot is indeed symmetrical with respect to the origin.

As a direct consequence of this theorem, we can state the following very useful property:[1] *The DP plot across two identical resistors connected back to back in series (Fig. 7-15b), or back to front in parallel (Fig. 7-15c) is symmetrical with respect to the*

---

[1] This theorem is valid for arbitrary, nonlinear resistors which satisfy the hypothesis. Hence, the DP plot and TC plot may be multivalued.

origin. Since the resistors need not be bilateral, this property is often used in practice for obtaining symmetrical DP plots from nonbilateral resistors. For example, the two–zener diode back-to-back circuit shown in Fig. 7-15*d* is widely used in practice to obtain a symmetrical zener v-i curve. Similarly, the tunnel-diode battery combination shown in Fig. 7-15*e* is used in many switching circuit applications (Chap. 16) which require a symmetrical DP plot.

Since the nonlinear lattice network is one of the most versatile and useful circuits in practice, let us take a further look at the circuit shown in Fig. 7-16*a*.[1] We shall state two general properties of this network.

SYMMETRIC LATTICE NETWORK: PROPERTY 1

If resistors R_1 and R_2 are identical and resistors R_3 and R_4 are identical, then corresponding to any value of v_2, the DP plot is symmetrical with respect to the origin.

This property can be easily verified by applying the v-shift theorem to obtain the equivalent circuit shown in Fig. 7-16*b*. If R_1 is identical with R_2, and R_3 is identical with R_4, then this circuit exhibits horizontal symmetry. The property, therefore, follows from the symmetric DP plot–TC plot theorem.

SYMMETRIC LATTICE NETWORK: PROPERTY 2

If the four resistors R_1, R_2, R_3, and R_4 in Fig. 7-16*c* are identical, then the v_o-vs.-v_i TC plot is symmetrical with respect to the v_o axis.

To prove this property, let us obtain the output-characteristic curve corresponding to each value $v_i = k$ of the circuit shown in Fig. 7-16*c*. If we apply the v-shift theorem, we obtain the equivalent network shown in Fig. 7-16*d*. Notice that since all four resistors are identical, the two black boxes N_1 and N_2 differ only in the polarity of the voltage source. If we multiply the terminal voltage v_i of both voltage sources by minus 1, the black box N_2 becomes black box N_1, and the black box N_1 becomes black box N_2. Hence, applying the series commutative property, we conclude that the DP plots across terminals *c-d* corresponding to $v_i = k$ and $v_i = -k$ are identical. This means that regardless of the load resistor R_L, the load-line construction must lead to a TC plot which is symmetrical with respect to the v_o axis.

---

[1] Observe that the polarities of the two resistors R_2 and R_4 are chosen opposite to those shown in Fig. 7-13*a*.

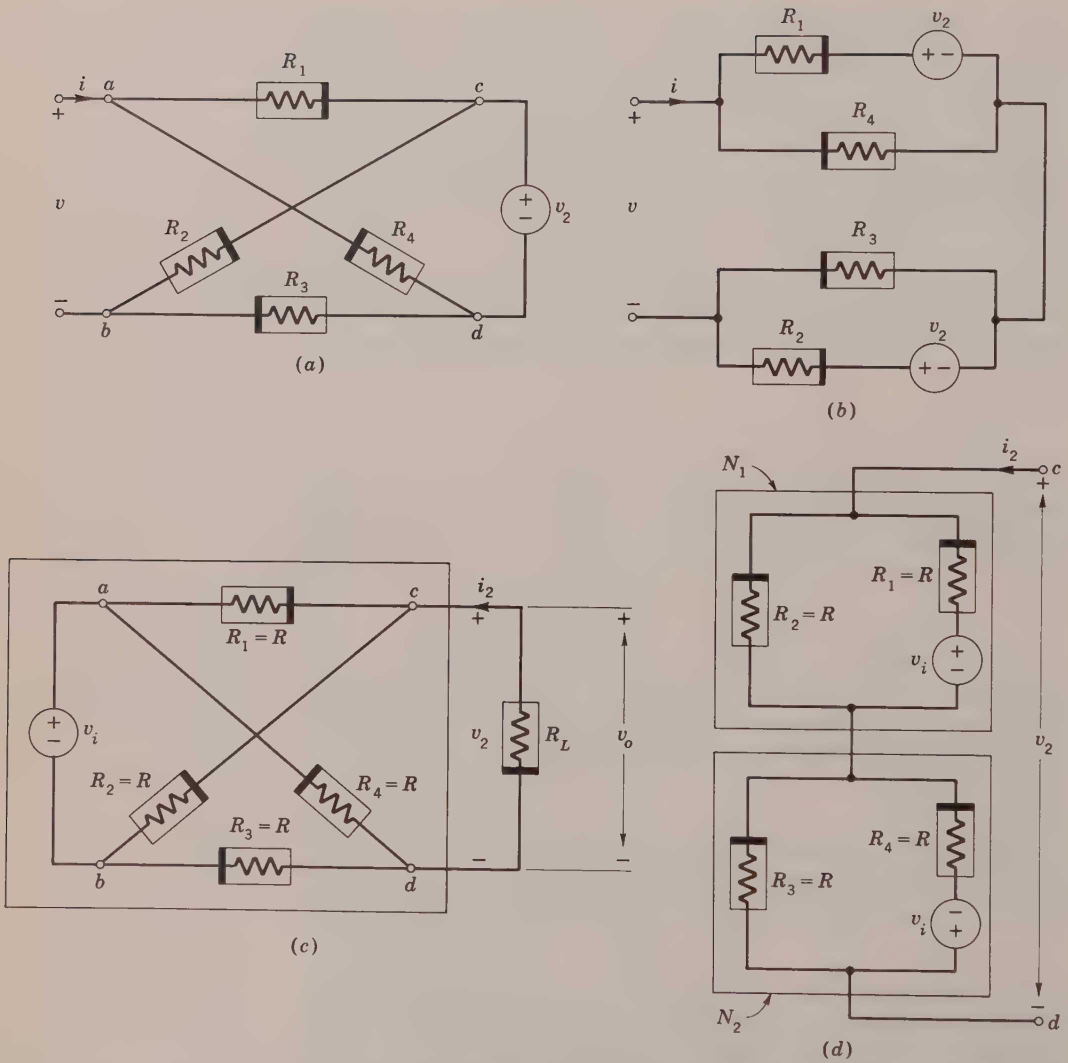


Fig. 7-16. Equivalent lattice networks for verifying the symmetric DP plot and TC plot properties of a symmetric nonlinear lattice.

In view of this property, we need to find only the TC plot corresponding to $v_i \geq 0$. The other half is simply a mirror reflection about the v_o axis. Since the full-wave rectifier circuit shown in Fig. 7-8*a* satisfies the hypothesis of this property, we could have saved half the work if we realized earlier that the TC plot is symmetrical.

Exercise 1: Give an example of a network which exhibits the following types of symmetry: (*a*) $\theta = 30$, 45, and 90° rotational symmetry. (*b*) Both flip-over and rotational symmetry.

Exercise 2: Show that a network exhibits $\theta = 180°$ rotational symmetry if, and only if, the network remains invariant after it is flipped over twice with respect to vertical and horizontal axes of symmetry.

Exercise 3: Prove the rotational symmetric-network theorem rigorously.

Exercise 4: Show that by flipping over one side of the network shown in Fig. 7-14*b*, the network is transformed into one which exhibits rotational symmetry. Verify that the same end result shown in Fig. 7-14*e* is obtained by this approach.

Exercise 5: Show that if all resistors in the lattice network of Fig. 7-13*a* are bilateral, then both the DP plot across terminals a-b and the v_o-vs.-v TC plot are symmetrical with respect to the origin.

7-8 COMPLEMENTARY NETWORKS

There are many practical networks which occur in pairs and therefore can be analyzed as a single entity. For example, two circuits which differ only in that one contains *p-n-p* transistors while the other contains *n-p-n* transistors should logically be analyzed as one circuit. In order to formulate a systematic procedure for doing this, we need the following definitions:[1]

DEFINITION OF COMPLEMENTARY RESISTORS

Given an n-terminal resistor R, we define an associated *complementary resistor* $\bar{R}$ by multiplying all voltage and current coordinates (including the parameters) of the characteristic curves of R by minus 1.

If $n = 2$, the $\bar{v}$-$\bar{i}$ curve of the complementary resistor $\bar{R}$ is obtained by rotating the v-i curve by 180° with respect to the origin. The complement of a bilateral resistor is therefore itself, and the complement of a nonbilateral resistor (including a voltage source and a current source) is obtained by transposing the terminals of the original resistor R.

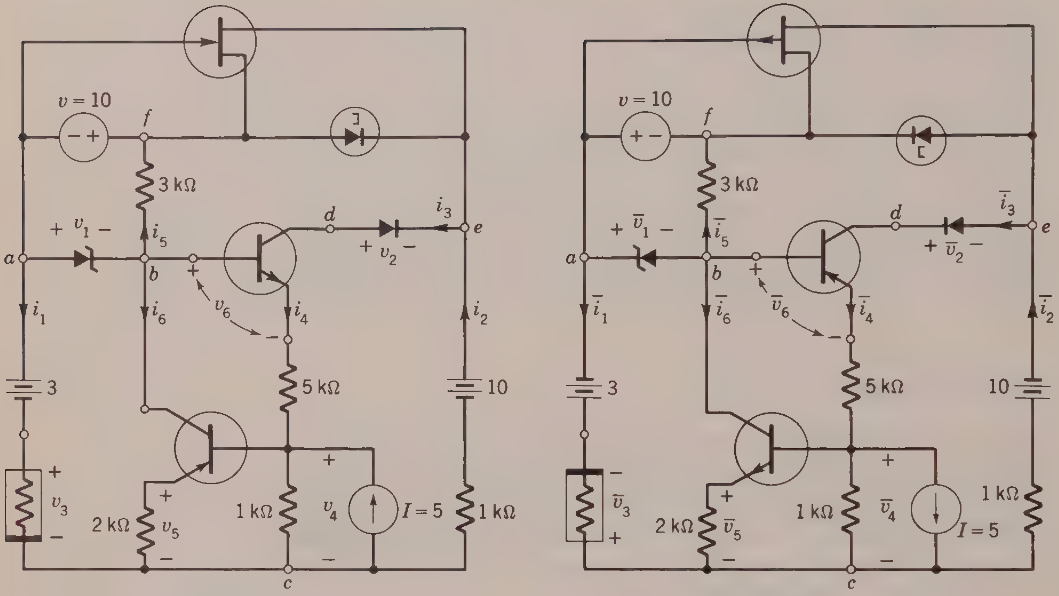
If we replace each resistor in a given network N by its complement, we obtain the *complementary network* $\bar{N}$. The following theorem is very useful for analyzing a large class of practical circuits.

COMPLEMENTARY-NETWORK THEOREM

If $\bar{N}$ is the complementary network associated with a given network N, then the voltage across any pair of nodes of $\bar{N}$ is equal to the negative of the corresponding voltage in N. Similarly, the cur-

---

[1] All definitions and concepts introduced in this section are applicable also to dynamic networks.



(a) (b)

Fig. 7-17. An example of two complementary nonlinear networks.

rent in any wire of $\overline{N}$ is equal to the negative of the corresponding current in N.

To prove this theorem, we first observe that the equations of motion of $\overline{N}$ and N are identical, provided we define each voltage and current variable of $\overline{N}$ to be the negative of the corresponding variable in N. The operating points of the two networks are therefore identical, apart from a negative sign.

This theorem is often used in practice for converting a network containing *p-n-p* transistors or *p*-channel FETs into a complementary network containing *n-p-n* transistors or *n*-channel FETs. This application is particularly useful in the design of "logic circuits" to be discussed in Chap. 10. As a simple example, consider the network shown in Fig. 7-17*a*. The complement of this network is easily constructed as shown in Fig. 7-17*b*. If the operating point of the original network is given by (v_j, i_j), $j = 1, 2, \ldots, 6$, then the corresponding operating point $(\bar{v}_j, \bar{i}_j)$ of the complementary network is given by $\bar{v}_j = -v_j$ and $\bar{i}_j = -i_j$, $j = 1, 2, \ldots, 6$.

Exercise 1: Show that if a network contains only passive bilateral resistors and no sources, then the voltage and current variables of this network must be identically zero.

Exercise 2: Explain how to obtain the complement of a controlled source.

Exercise 3: Find the complement of a gyrator, a rotator, a scalor, a reflector, and an ideal transformer.

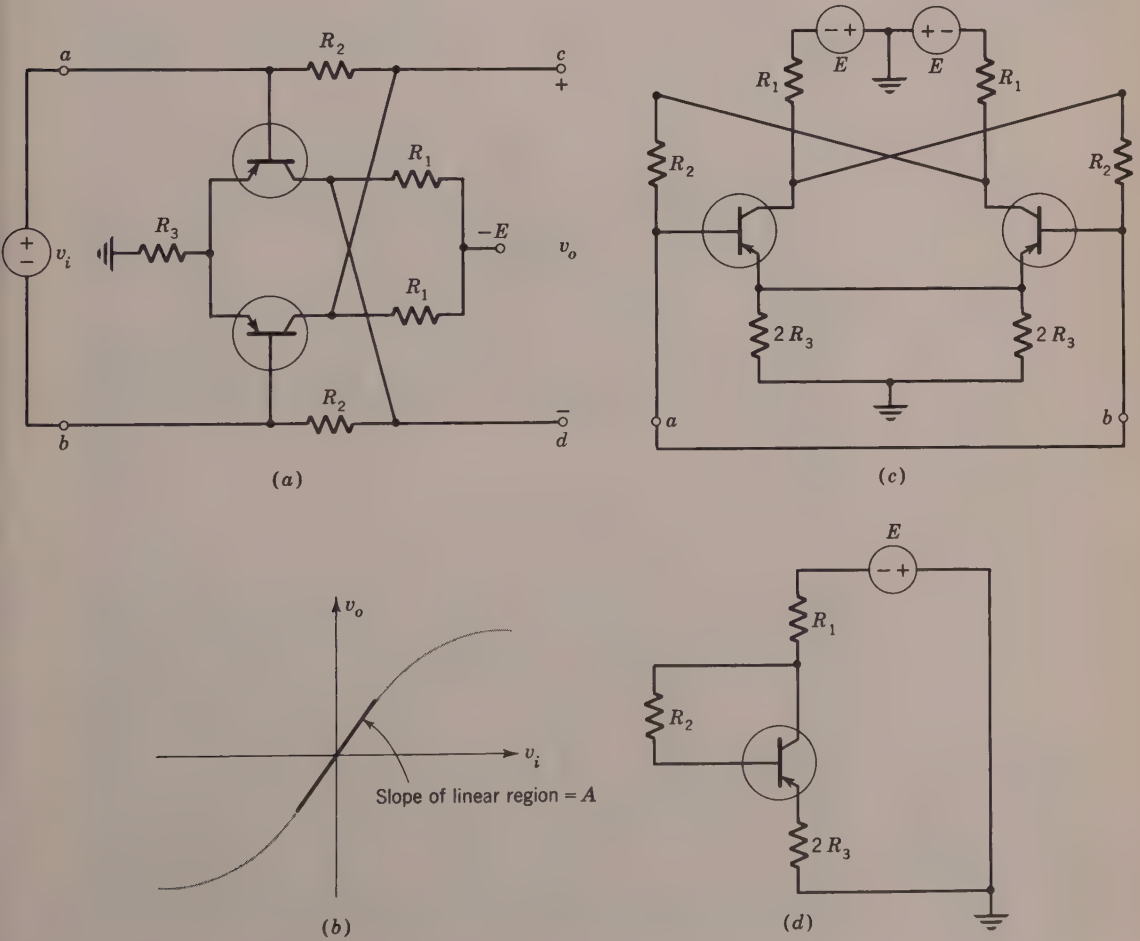
7-9 SOME PRACTICAL CIRCUITS

We shall now show how the principles introduced so far can be used to analyze many practical circuits of interest.

7-9-1 DC AMPLIFIER

The circuit shown in Fig. 7-18*a* can be designed to give a TC plot $v_o = Av_i$ over some range of input voltage v_i. Since this circuit is usually used with a dc input voltage and since $A \gg 1$, it is called a

Fig. 7-18. An illustration of the application of symmetry principles in simplifying the analysis of a dc amplifier circuit.



dc amplifier. The black box seen by the driving-point voltage source v_i exhibits horizontal flip-over symmetry. Hence, both the DP plot and the v_o-vs.-v_i TC plot are symmetrical with respect to the origin. A typical TC plot is shown in Fig. 7-18*b*.

The first step in the design of a dc amplifier is the determination of the operating point when $v_i = 0$. If we redraw this circuit as shown in Fig. 7-18*c*, we find that it exhibits vertical flip-over symmetry.[1] Under appropriate design, this circuit has a unique solution, and, hence, applying the flip-over symmetric-network theorem, we may open-circuit the horizontal wires and short-circuit the crossed wires and obtain the equivalent circuit (only one-half is shown) shown in Fig. 7-18*d*. The operating point of this much-simplified circuit can now be easily obtained.

There are several practical reasons for preferring symmetrical circuits over unsymmetrical ones. The most important reason is that the output voltage v_o of symmetric circuits such as the dc amplifier shown in Fig. 7-18*a* is not sensitive to changes in element parameter values, provided the changes take place also in a symmetrical sense. For example, if $v_i = 0$ in Fig. 7-18*a*, then $v_o = 0$ regardless of the values of E, R_1, R_2, and R_3, provided that each of the two symmetrically located resistors changes by an equal amount. Since all practical circuits are subject to variation due to temperature, aging, supply voltage changes, etc., the use of symmetrical circuits tends to minimize the effect resulting from such changes. This principle is widely used in the fabrication of integrated circuits because it is known that the circuit parameters of an integrated circuit vary in an amazingly uniform and symmetrical manner.

7-9-2 COMPLEMENTARY SYMMETRIC AMPLIFIER

The circuit shown in Fig. 7-19*a* contains an *n-p-n* transistor T_1 and a *p-n-p* transistor T_2 and is called a *complementary symmetric amplifier* because the network graph is symmetrical and the transistors are complementary. In order to obtain the v_o-vs.-v_i TC plot graphically, we apply the v-shift theorem and obtain the equivalent network shown in Fig. 7-19*b*. Notice that this circuit can be considered as a load resistor R across the parallel combination of the two black boxes N_1 and N_2. It suffices, therefore, to determine the DP plot across terminals *a-b* corresponding to each value $v_i = k$. Observe that we can save half of the work by taking advantage of the fact that the network N_1 with $v_i = k$ is the complement of the network N_2 with $v_i = -k$, and vice versa. Hence, we have to find

---

[1] It is possible for this circuit to have more than one solution. Hence, we must watch out for the possibility of extraneous, and hence erroneous, solutions.

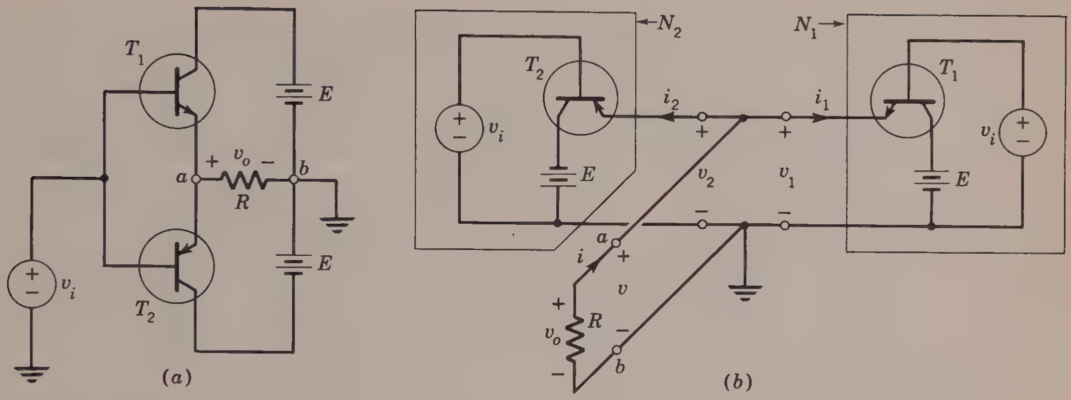


Fig. 7-19. An example of a complementary symmetric amplifier circuit.

only the system of v_1-i_1 curves corresponding to both positive and negative values of v_i in order to determine the DP plot across terminals *a-b*.

7-9-3 PUSH-PULL AMPLIFIER

We shall conclude this section with a discussion of one of the most commonly used nonlinear networks, the push-pull amplifier circuit shown in Fig. 7-20*a*. The input is excited by two *identical* voltage sources v_i, and the output voltage is taken across the collectors of the two transistors. With the exception of the polarity of the voltage source v_i, this circuit exhibits a horizontal flip-over symmetry. Hence, by the same argument as before, we conclude that the TC plot between v_o and v_i must be symmetrical with respect to the origin. To actually determine the TC plot, we apply the *v*-shift theorem to batteries E_1 and E_2 and redraw the network into the more convenient form shown in Fig. 7-20*b*. Since the wire connecting the upper and lower portions of the circuit is a cut set by itself, the current in the wire is zero, and hence we may cut the wire without affecting the solution of the network. The TC plot from the collector-emitter terminals to the input voltage source can be easily obtained by the double load-line method. The typical TC plots for v_{ab} versus v_i and $-v_{cd}$ versus v_i are shown, respectively, in Fig. 7-20*c* and *d*. An inspection of Fig. 7-20*b* shows that $v_o = v_{ab} - v_{cd}$; hence, the desired v_o-vs.-v_i TC plot is simply obtained by addition of the TC plots shown in Fig. 7-20*c* and *d*. The resulting plot shown in Fig. 7-20*e* is indeed symmetrical with respect to the origin.

In order to appreciate why this circuit is useful, the output voltage waveforms for $v_{ab}(t)$, $-v_{cd}(t)$, and $v_o(t)$ are shown in Fig.

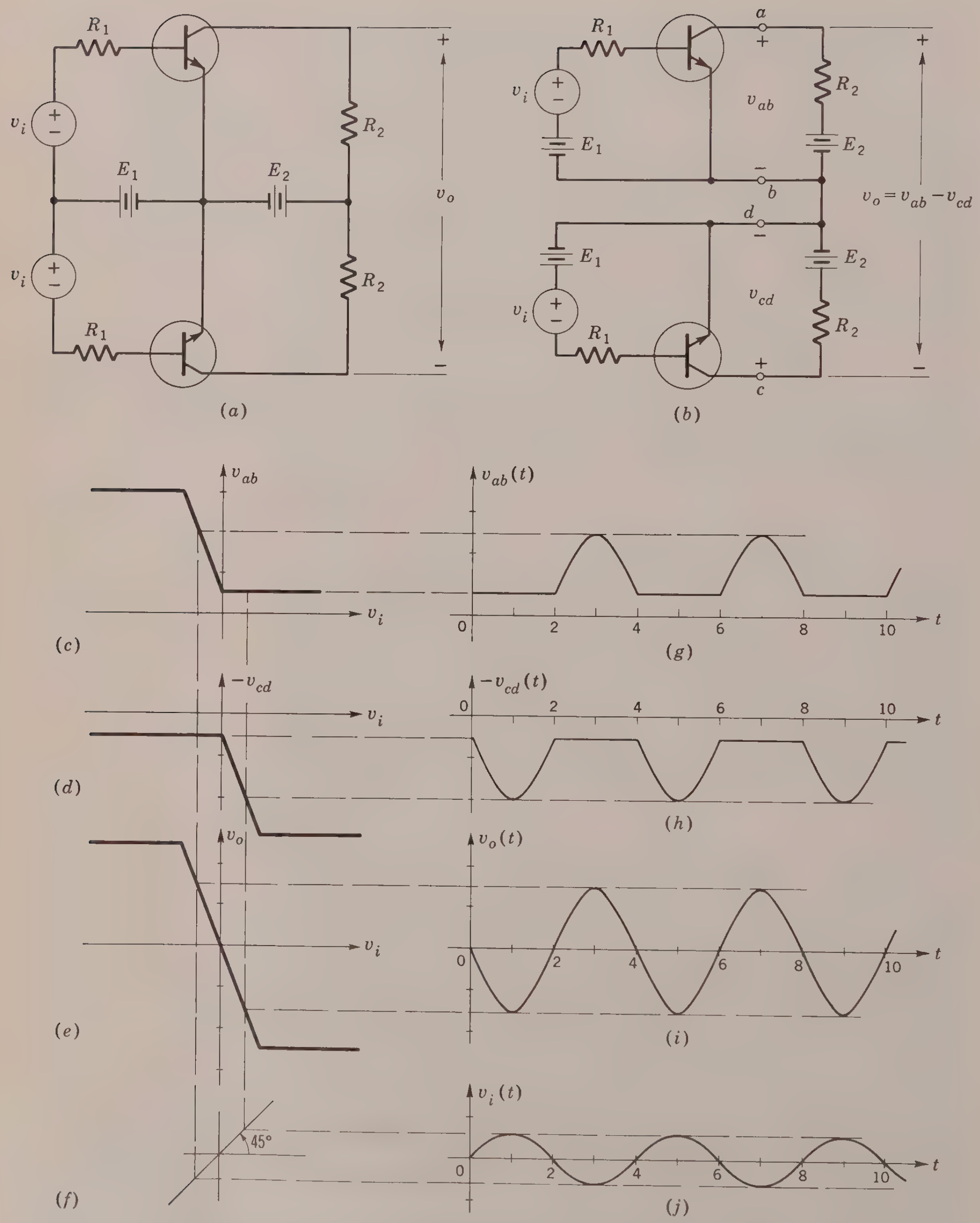


Fig. 7-20. The graphical procedure for determining the TC plot of a push-pull amplifier circuit.

7-20g to *i*, respectively, corresponding to the input sinusoidal voltage $v_i(t)$ shown in Fig. 7-20*j*. Notice that both $v_{ab}(t)$ and $v_{cd}(t)$ suffer from serious distortion every half-cycle. Combining these signals, however, tends to cancel the distortion, so to speak, thereby resulting in a high-voltage amplified sinusoidal waveform of high fidelity. In view of this property, most good hi-fi sets use a push-pull amplifier in the output stage.

In practice, the two identical voltage sources v_i are usually replaced by one voltage source and a three-winding transformer as shown in Fig. 7-21*a*. This transformer is a practical model of the three-port ideal transformer defined earlier by Eq. (3-47) of Sec. 3-5-2. For convenience, let us rewrite this equation with $n_2 = n_3$; thus

$$v_2 = \frac{n_2}{n_1} v_1 \tag{7-1}$$

$$v_3 = \frac{n_2}{n_1} v_1 \tag{7-2}$$

$$n_1 i_1 + n_2 i_2 + n_2 i_3 = 0 \tag{7-3}$$

In order to understand the action of this transformer, let us show that it is equivalent to the network N_2 shown in Fig. 7-21*b*. Writing the KVL and KCL equations for N_1, we obtain

$$v_1 = v_s(t) - Ri_1 \tag{7-4}$$

$$i_2 = I_1 \tag{7-5}$$

$$i_3 = I_1 + I_2 \tag{7-6}$$

$$E_1 = v_2 + v_3 \tag{7-7}$$

$$E_2 = v_3 \tag{7-8}$$

Substituting Eq. (7-3) for i_1 in Eq. (7-4), we obtain

$$v_1 = v_s(t) + R\frac{n_2}{n_1}(2I_1 + I_2) \tag{7-9}$$

Substituting Eqs. (7-1), (7-2), and (7-9) into Eqs. (7-7) and (7-8), we obtain

$$E_1 = 4\left(\frac{n_2}{n_1}\right)^2 RI_1 + 2\left(\frac{n_2}{n_1}\right)^2 RI_2 + 2\left(\frac{n_2}{n_1}\right) v_s(t) \tag{7-10}$$

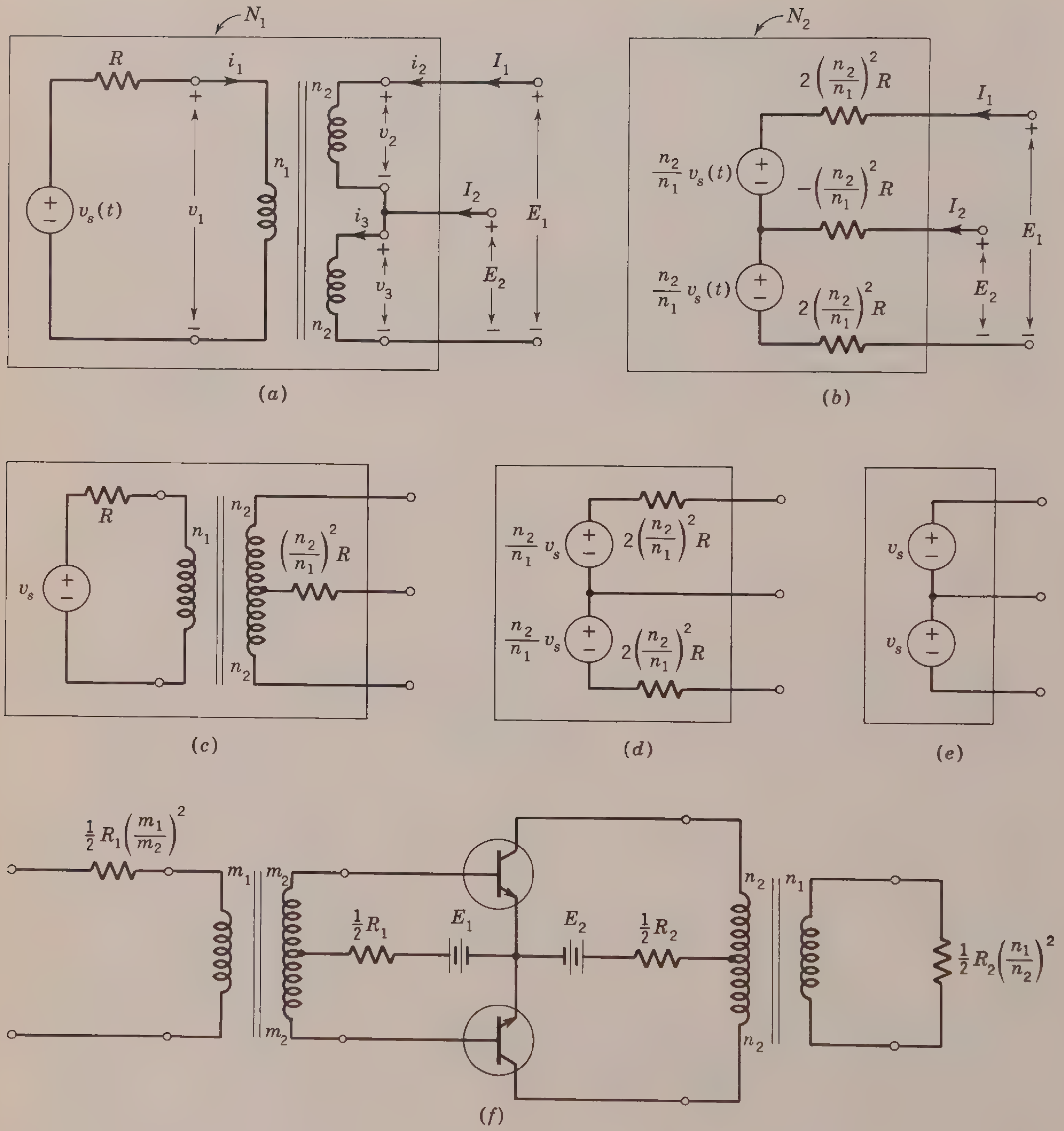


Fig. 7-21. An illustration of the use of a three-port ideal transformer in place of two back-to-back voltage sources required by the push-pull amplifier.

$$E_2 = 2\left(\frac{n_2}{n_1}\right)^2 RI_1 + \left(\frac{n_2}{n_1}\right)^2 RI_2 + \left(\frac{n_2}{n_1}\right) v_s(t) \tag{7-11}$$

But Eqs. (7-10) and (7-11) are exactly the equations characterizing the network N_2 of Fig. 7-21*b*, and hence N_1 and N_2 are indeed equivalent.

Notice that if we add a resistance $(n_2/n_1)^2R$ in series with the center wire as shown in Fig. 7-21*c*, the negative resistance drops out as shown in Fig. 7-21*d*. Notice also that this circuit can be further simplified to that shown in Fig. 7-21*e* by setting $R = 0$ and $n_2 = n_1$ in Fig. 7-21*a*. Hence, we have demonstrated that a three-winding transformer can be substituted in place of two identical voltage sources connected back to back in series. This technique is widely used in practice. In fact, two of these transformers are commonly used in place of the voltage source v_i and resistors R_2 of the push-pull amplifier as shown in Fig. 7-21*f*. The values of the resistances have been chosen to show that the circuits in Figs. 7-20*a* and 7-21*f* are exactly equivalent. In practice, this is of course unnecessary. In fact, we can save the battery E_1 by shorting it and connecting a resistor from the transformer center tap to the positive side of the battery E_2.

Exercise 1: Explain why we split the voltage source E and resistor R_3 into two identical parts as shown in the equivalent network in Fig. 7-18*c*.

Exercise 2: Give the step-by-step procedure for determining the TC plot of the complementary symmetric amplifier circuit in Fig. 7-19.

Exercise 3: Find the TC plots shown in Fig. 7-20 for the following set of specifications: $R_1 = 500\ \Omega$, $R_2 = 200\ \Omega$, $E_1 = 0.5$ volt, $E_2 = 10$ volts, and assuming the two transistors to be type 2NXX17A (see Appendix D).

Exercise 4: Find the equivalent circuit across port 2 of a two-port ideal transformer with port 1 connected to a voltage source v_s and resistor R in series.

Exercise 5: (*a*) Verify that the push-pull amplifier circuits shown in Figs. 7-20*a* and 7-21*f* are exactly equivalent. (*b*) Show that the battery E_1 can be traded by a resistor R of appropriate resistance. (*c*) Specify the value of R in terms of the voltage E_1.

7-10 SUMMARY

Equivalent networks Two n-terminal networks N_1 and N_2 are equivalent if, and only if, they can be interchangeably connected to *any* external network N without affecting the solution of the elements inside N. It is meaningless to compare the solutions inside N_1 and N_2.

Some useful equivalent network properties

1. Series (parallel) commutative property
2. Redundant series (shunt) element property
3. Equivalent current (voltage) source property

Some useful equivalent network theorems

1. Nonlinear Thévenin (Norton's) theorem
2. Voltage-source substitution theorem
3. Current-source substitution theorem
4. Short-circuit substitution theorem
5. Open-circuit substitution theorem
6. *v*-shift theorem
7. *i*-shift theorem

Symmetry

1. Flip-over Symmetry: Axis of symmetry lies on the plane of the network. Common types are vertical and horizontal symmetry.
2. Rotational Symmetry: Axis of symmetry is perpendicular to the plane of the network. Common type is 180° rotational symmetry.

Some useful theorems on symmetry

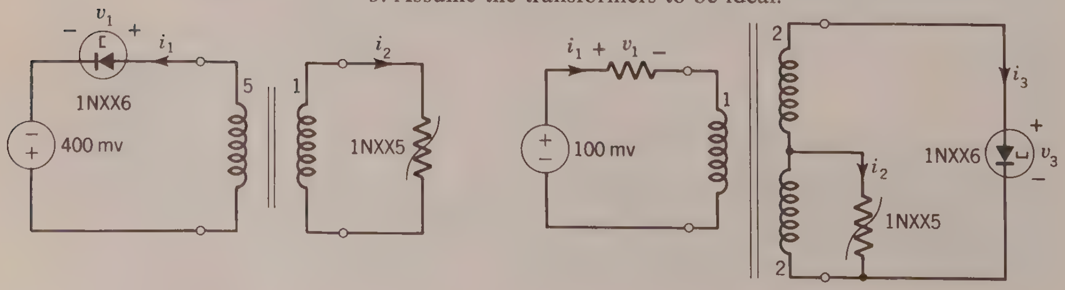
1. Flip-over symmetric-network theorem
2. Rotational symmetric-network theorem
3. Symmetric lattice network theorem
4. Symmetric DP plot–TC plot theorem

Complementary-network theorem The voltage and current solution of a given network N is equal to the negative of the complementary network $\overline{N}$.

PROBLEMS

7-1 Find the operating points of the networks shown in Fig. P7-1*a* and *b*. Assume the transformers to be ideal.

Fig. P7-1.



(*a*) (*b*)

7-2 The network shown in Fig. P7-2 contains two nonmonotonic elements.

(a) Find the operating points of the network.

(b) Apply the vertical flip-over symmetry theorem to the network and find the operating points. Is the result the same as in part (*a*)? Why or why not?

(c) Change the resistor from 500 to 100 Ω and repeat parts (*a*) and (*b*).

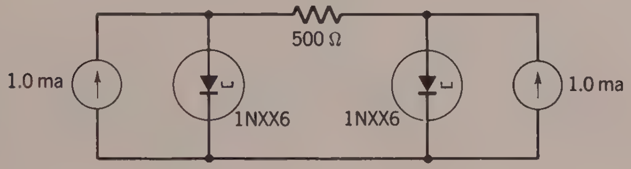


Fig. P7-2.

7-3 The nonlinear lattice network in Fig. P7-3 contains two 1NXX7 back diodes, which are nonmonotonic elements.

(a) Use graphical analysis to find the operating points of the network. (HINT: Apply the *v*-shift theorem.) Is the operating point unique?

(b) Apply the vertical flip-over symmetric-network theorem to the network and find the operating points.

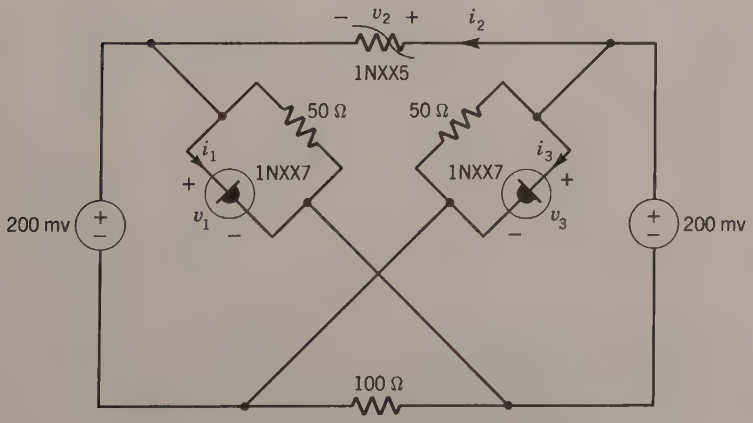


Fig. P7-3.

7-4 The network in Fig. P7-4 contains three type 1NXX5 varistors.

(a) Use graphical analysis to find the operating points of the network.

(b) The network has rotational symmetry. Use this property to find the operating points. Check your answer with part (*a*).

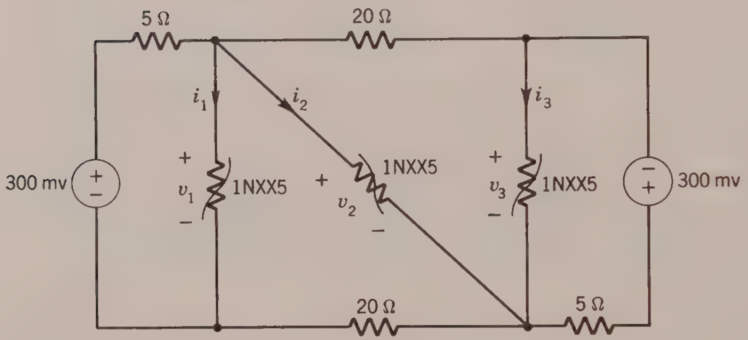


Fig. P7-4.

7-5 For the network shown in Fig. P7-5, apply the *v*-shift theorem to find the following:

(a) The voltage across the varistor when $v = 100$ mv

(b) The DP plot across terminals *a*-*b*

(c) The v_o-vs.-v TC plot

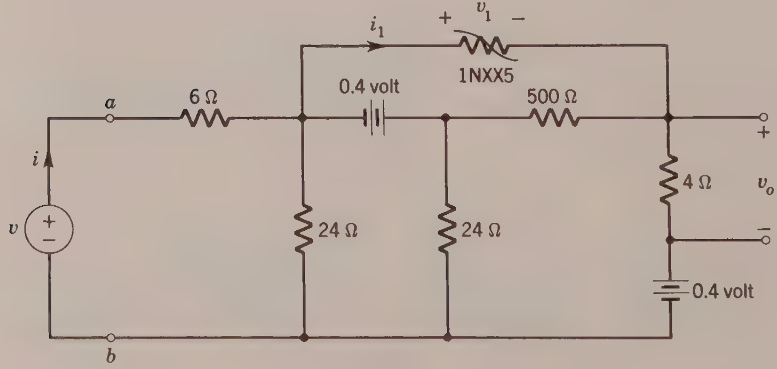


Fig. P7-5.

7-6 For the network shown in Fig. P7-6, apply the *i*-shift theorem to find the following:

(a) The voltage across the varistor when $i = 100$ ma

(b) The DP plot across terminals *a*-*b*

(c) The v_o-vs.-i TC plot.

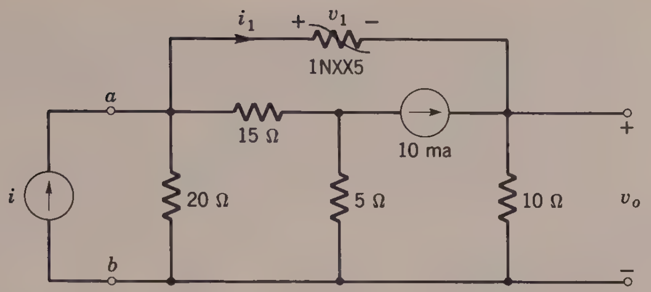


Fig. P7-6.

7-7 Consider the symmetric nonlinear lattice network shown in Fig. 7-13*a* of the text. Find the operating points of the network corresponding to the v-i curves for R_A, R_B, and R_L in Fig. P7-7*a* to *c*. Assume that $v = 10$ volts. Use the symmetric lattice network theorem to verify your answers.

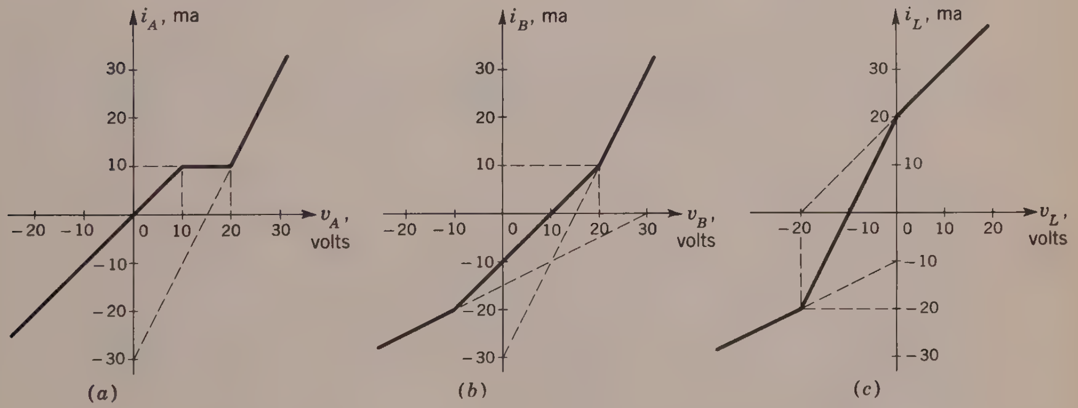
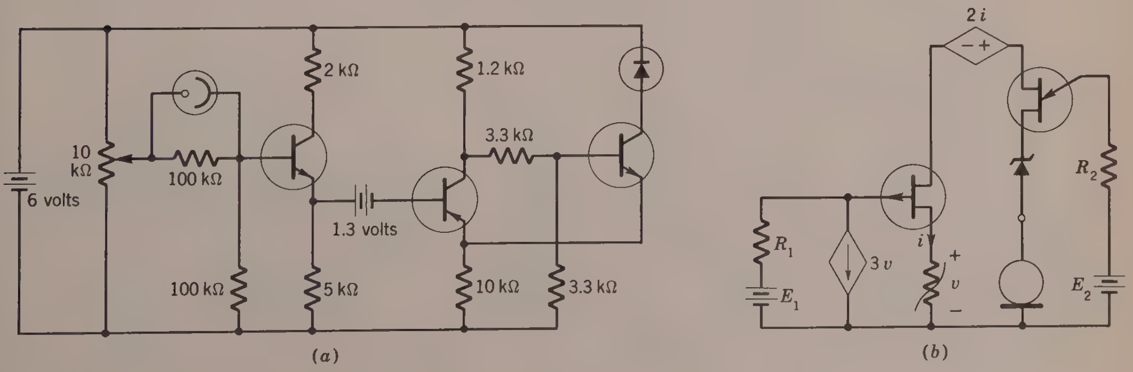


Fig. P7-7.

7-8 Find the complements of the networks shown in Fig. P7-8*a* and *b*.

Fig. P7-8.



7-9 The two basic circuits shown in Fig. P7-9*a* and *b* are usually used in practice to detect changes in temperature and light illumination.

(a) Give the step-by-step procedure for determining the TC plots of these circuits.

(b) Give the principles of operation.

(c) Why is a bridge circuit chosen for this purpose?

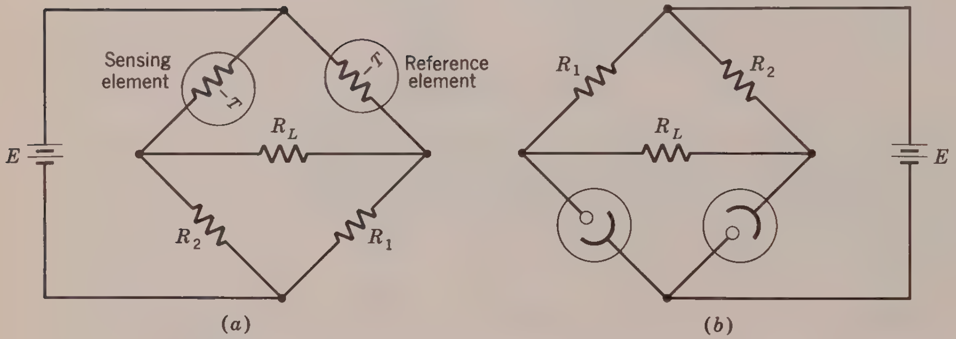


Fig. P7-9.

7-10 Give the step-by-step procedure for determining the characteristic curves of the three common compound transistor circuits shown in Fig. P7-10*a* to *c*.

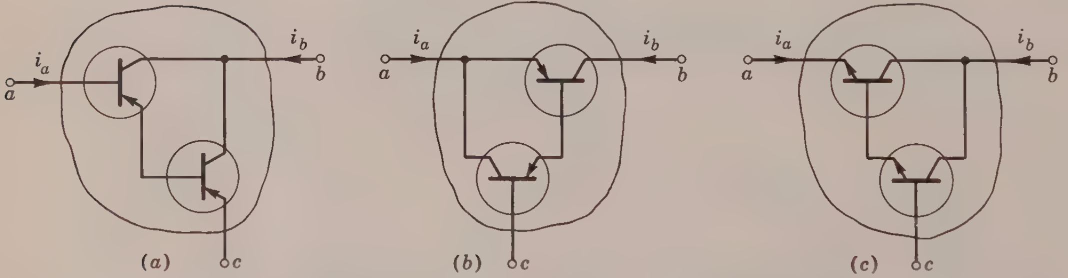
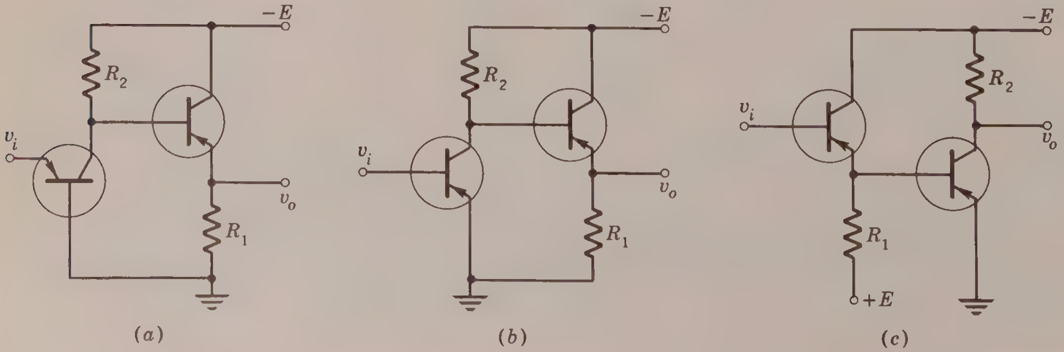


Fig. P7-10.

7-11 Give the step-by-step procedure for determining the v_o-vs.-v_i TC plots of the three common direct-coupled amplifier circuits shown in Fig. P7-11*a* to *c*.

Fig. P7-11.



7-12 Give the step-by-step procedure for determining the operating point of the networks shown in Fig. P7-12*a* and *b* when $v_i = 0$. Do the DP plot and TC plot of these networks possess any form of symmetry?

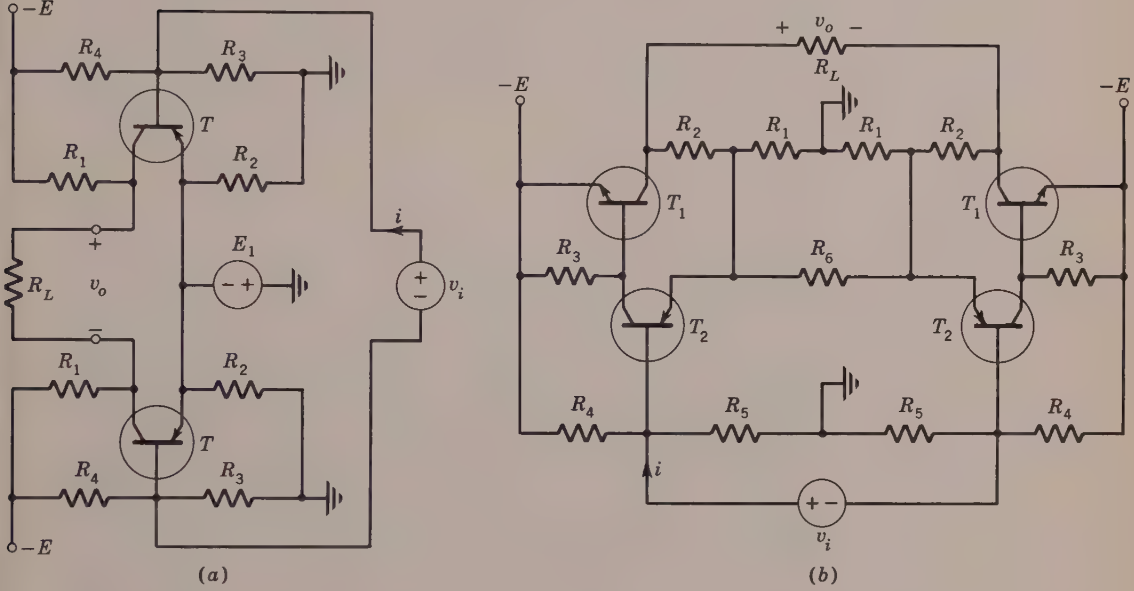
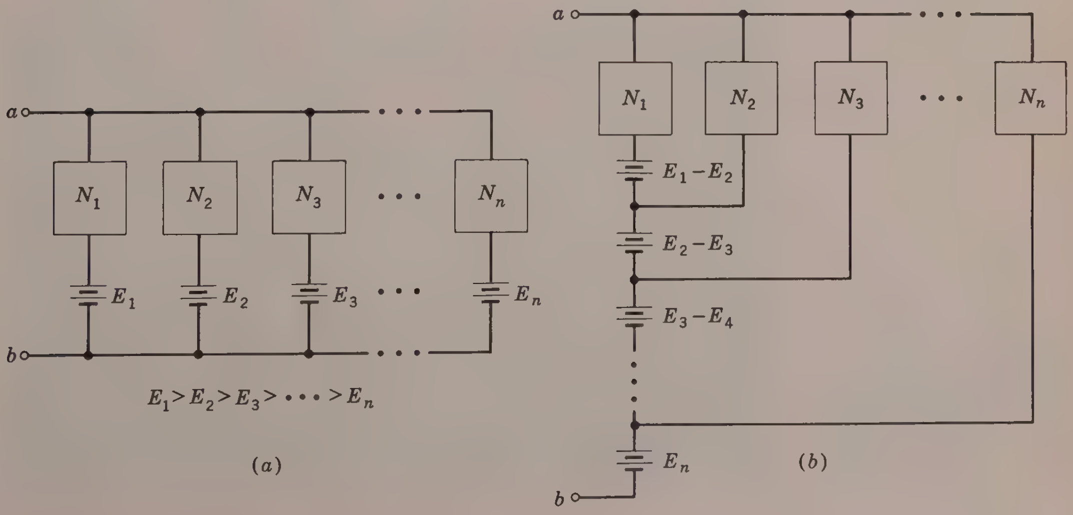


Fig. P7-12.

7-13 Prove that the two networks shown in Fig. P7-13*a* and *b* are equivalent, thus allowing the use of one multi-tap battery instead of *n* batteries.

Fig. P7-13.



8 GRAPHICAL SYNTHESIS OF RESISTIVE NONLINEAR NETWORKS

8-1 WHAT IS NETWORK SYNTHESIS?

So far, we have been doing "network analysis," and the emphasis has been on the determination of the operating points, the DP plots, and the TC plots of resistive nonlinear networks. Consider now the converse problem: Given a set of specifications in the form of an operating point, a DP plot, a TC plot, or a combination of them, find a network with the prescribed specifications. The study of this problem is called *network synthesis.*

There is a basic difference between network analysis and network synthesis. In network analysis, there is always an answer. However, in network synthesis, an answer does not always exist. It is possible that no network can be found to have a certain specification, in which case the specification is said to be *unrealizable.* Another basic difference between analysis and synthesis is the fact that if a given specification is realizable, then there are generally many distinct networks that can be found which have the same specifications.

The possibility of more than one realization suggests that a particular realization may be better than another with respect to some practical criteria such as complexity, economy, reliability, sensitivity, weight, and space factors. For example, one realization may be better than another with respect to economy but may be inferior to another with respect to reliability. The point we are driving at is that there is no such thing as the best realization per se. However, in any practical problem, there is always an optimum choice with respect to some criteria. The task of choosing the optimum realization from among a number of possibilities is called *network design.*

It is important to distinguish between network synthesis and network design. In network synthesis we are interested only in finding a network that meets the specifications. No other consideration is involved. In network design we go one step further and choose an optimum realization *with respect to some design criteria* which often involve economic and other nonelectrical considerations. Hence, much personal judgment and experience are often necessary to arrive at an optimum realization. In view of this subjective nature of network design, only network synthesis will be considered in this book.

8-2 SYNTHESIS OF OPERATING POINTS

The first step in the design of many electronic circuits is to synthesize a network so that when it is connected to a two-terminal or three-terminal resistor R, a prescribed operating point is realized. The network for achieving this task is called a *biasing network.* There are many reasons why a particular operating point may be preferred over another. In the case of an amplifier, an optimum operating point might be one which leads to more gain (amplification of the input signal) and less distortion (departure from the desired signal). In the case of switching circuits, a number of suitably spaced operating points may be prescribed in order to carry out some logic function. Other considerations may involve power dissipation and sensitivity of the operating point to temperature variations.

8-2-1 BIASING A TWO-TERMINAL NONLINEAR RESISTOR

Given a two-terminal resistor R with a v-i curve Γ, let us study some methods for synthesizing a biasing network N as shown in Fig. 8-1*a* in order that resistor R will operate at some prescribed point Q on Γ. If Γ is either voltage-controlled or current-controlled, then the simple biasing network shown in Fig. 8-1*b* can always be chosen so that its load line passes through Q. For example, the load-line construction in Fig. 8-1*c* illustrates how a tunnel diode can be biased. Notice that this biasing network is not unique because we can choose many different values for the battery voltage E and solve for the corresponding resistance R_L, or vice versa.[1]

All practical devices are temperature-sensitive to some extent. Over a given range of temperatures $T_1 < T < T_2$, the v-i curve of the tunnel diode may vary within the band shown in Fig. 8-1*d*. Accordingly, the operating point will also change with temperature. In some sensitive instruments, it is usually important that the

---

[1] In practice, however, we are usually constrained to use standard battery voltages. In this case, the biasing network is unique with respect to the prescribed voltage.

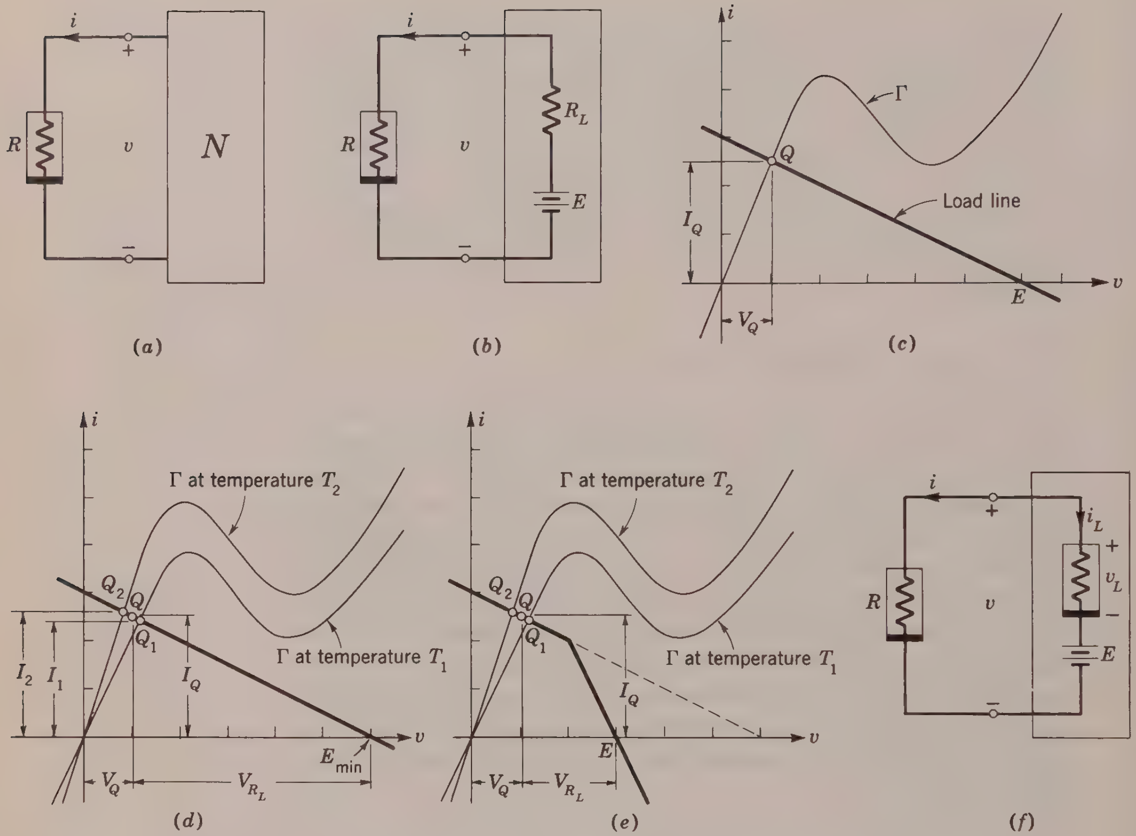


Fig. 8-1. An example of the synthesis of a linear and a nonlinear biasing network for a two-terminal resistor.

operating current I_Q lie between two prescribed bounds, such as $I_1 < I_Q < I_2$ corresponding to the point Q_1 at T_1 and the point Q_2 at T_2. In this case, it is necessary that $E \geq E_{\min}$, where $E_{\min}$ is the voltage intercept of the load line through the points Q_1 and Q_2 in Fig. 8-1*d*.

An examination of Fig. 8-1*d* shows that the power dissipated in the biasing resistor R_L is equal to $p = V_{R_L} I_Q$, where $V_{R_L} = E_{\min} - V_Q$. Since this power is merely wasted as heat, there are applications, such as in a spacecraft where power must be conserved, which dictate the use of a nonlinear biasing network. For example, a nonlinear load line may be chosen so that it passes through the points Q_1 and Q_2 as shown in Fig. 8-1*e*. Since the voltage V_{R_L} across the biasing resistor is now smaller, we have achieved the same objective but with less power dissipation. The nonlinear load line in this example can be realized by the passive nonlinear resistor-battery network shown in Fig. 8-1*f*. The passive nonlinear resistor can be realized by synthesizing an appropriate DP plot by the methods to be presented shortly.

No discussion of biasing techniques is complete without a consideration of the *sensitivity* of the operating point as a function of temperature. Perhaps a specific example will illustrate some of the practical problems which we have not yet considered.

EXAMPLE

The *v*-*i* curve of a typical nonlinear resistor is shown in Fig. 8-2*a* as a function of the ambient temperature. Assuming that the resistor has a maximum allowable power dissipation of 80 mw, it is desired to design a biasing network so that (1) at $T = 25°\text{C}$ the

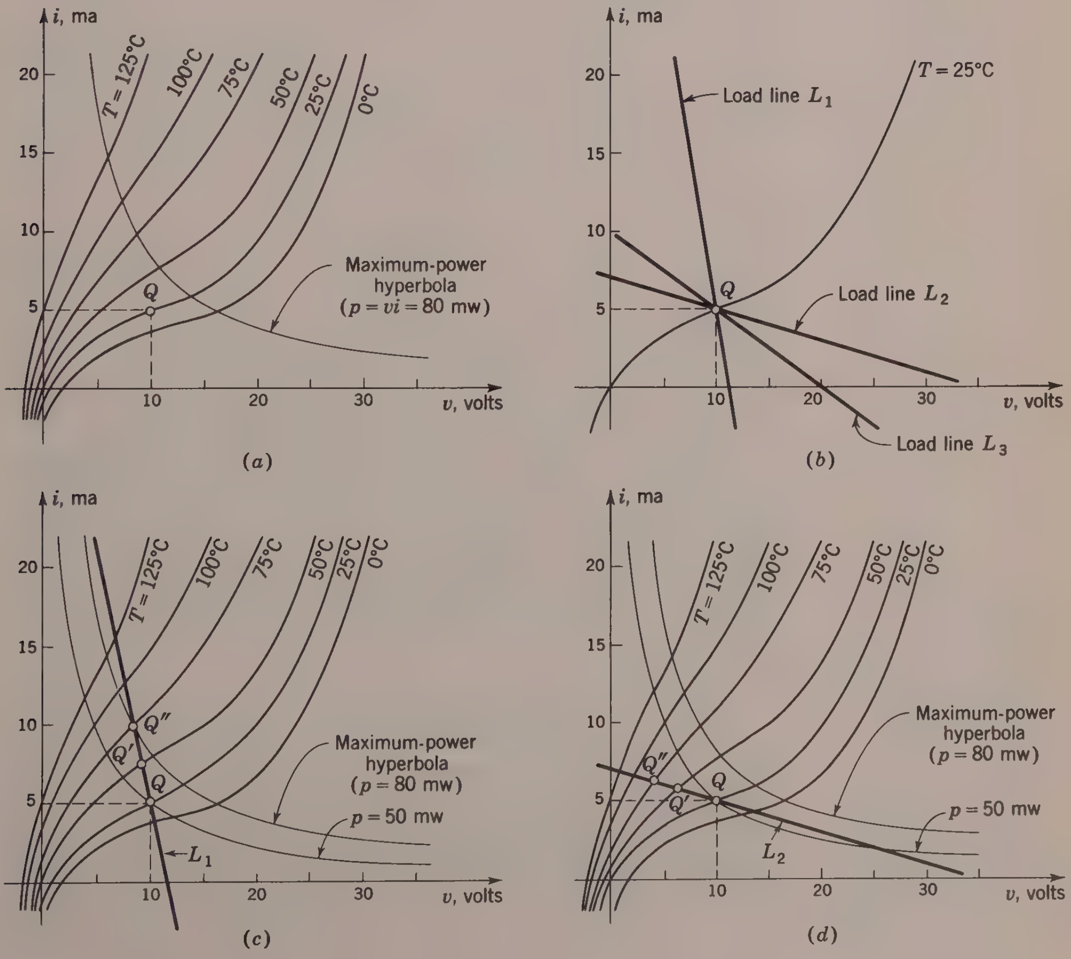


Fig. 8-2. An example of the thermal-runaway phenomenon in a two-terminal resistor.

operating point is at Q, and (2) the resistor is to be used in an environment where the temperature ranges from 25 to 150°C. This condition requires that the operating point be located below the maximum-power hyperbola $p = vi = 80$ mw at all times.

There are infinitely many load lines that will intersect the v-i curve (corresponding to $T = 25°$) at Q. Three typical load lines are shown in Fig. 8-2*b*. However, not all these load lines will satisfy condition 2.

Consider first the load line L_1. Let us draw the constant-power hyperbola passing through Q, namely, $p = 5(10) = 50$ mw, as shown in Fig. 8-2*c*. Now observe that as the temperature increases to $T = 50°$, the operating point shifts to Q'. Although Q' is still below the maximum-power hyperbola, it is above the constant-power hyperbola passing through Q. This means that the power dissipation at point Q' is greater than the power dissipation at point Q. If allowed to persist, this increase in power dissipation might produce enough heat to raise the surrounding temperature. For example, the temperature might eventually reach $T = 75°$. At this temperature, the operating point shifts to Q''. Observe that Q'' lies just below the maximum-power hyperbola. If this new operating point is allowed to persist, the surrounding temperature might again increase. The increase in temperature will shift the operating point still closer toward the maximum-power hyperbola. Since the process is cumulative and unstable, the operating point will eventually cross the maximum-power hyperbola, thereby destroying the device. This unstable phenomenon which leads to the eventual destruction of the device is usually called *thermal runaway*.

Before we determine the condition to avoid thermal runaway, let us consider the load line L_2. From Fig. 8-2*d* we observe that as the temperature increases from 25 to 50°C, the operating point shifts from Q to Q'. This time, however, Q' lies below the constant-power hyperbola passing through Q. This means that the power dissipation at point Q' is even less than the power dissipation at point Q. Therefore, the temperature will not increase, and the operating point is stable.

It is easy to see that there must be a critical load line between L_1 and L_2 which separates the unstable thermal-runaway phenomenon from the stable one. In fact, we observe from Fig. 8-2*d* that the maximum allowable slope of the load line passing through Q must be equal to the tangent at Q of the $p = 50$ constant-power hyperbola. Any slope greater than the tangent at Q would lead to a destructive thermal-runaway phenomenon.

Since the thermal-runaway phenomenon is an important problem to prevent in practice, let us derive a general criterion. For convenience, the positive-slope portion of the v-i curve of a two-terminal resistor is said to exhibit a *negative temperature coefficient*[1] if $\partial v/\partial T < 0$. This means that for each fixed value of i, the voltage v of the v-i curve decreases as the temperature T increases. Similarly, the positive-slope portion of the v-i curve is said to exhibit a *positive temperature coefficient* if $\partial v/\partial T > 0$. From the above consideration, it is clear that to prevent thermal runaway of a negative temperature coefficient resistor, the *minimum* value of the load resistor R_L must be such that the load line is tangent to the constant-power hyperbola $vi = p_Q$ at the operating point (I_Q,V_Q). The slope at Q is given by

$$\left.\frac{\partial i}{\partial v}\right|_Q = \left.\frac{-p_Q}{v^2}\right|_{v=V_Q} = \frac{-p_Q}{V_Q{}^2} \tag{8-1}$$

But $p_Q = V_Q I_Q$ at Q, hence Eq. (8-1) becomes

$$\left.\frac{\partial i}{\partial v}\right|_Q = -\frac{I_Q}{V_Q} \tag{8-2}$$

From Eq. (8-2) we obtain the criterion $R_L > V_Q/I_Q$. Similarly, to prevent thermal runaway in a positive temperature coefficient resistor, it is easily seen that the criterion is $R_L < V_Q/I_Q$.

The preceding discussion has neglected several factors such as the heat transfer by conduction, convection, and radiation from the resistor to the outside world. These thermal factors must be considered in any practical design. Since they are mostly empirical in nature, they are outside the scope of this book. From a purely circuit point of view, however, which is often rather conservative, we can summarize the above results in the form of a theorem.[2]

THERMAL-RUNAWAY CRITERION FOR TWO-TERMINAL RESISTORS

Let (I_Q,V_Q) be the prescribed operating point of a two-terminal resistor. In order to prevent the thermal-runaway phenomenon from occurring, it is sufficient that the load resistance R_L be greater than V_Q/I_Q if the resistor has a negative temperature coefficient, and be less than V_Q/I_Q if the resistor has a positive temperature coefficient.

---

[1] This is consistent with the popular usage where a temperature coefficient is said to be negative if the dc resistance corresponding to the *linear* portion of the v-i curve decreases as the temperature increases.

[2] Since conduction, convection, and radiation all work in our favor, the criterion in this theorem should be interpreted as a *worst case* criterion.

8-2-2 BIASING A THREE-TERMINAL NONLINEAR RESISTOR

Given the input- and output-characteristic curves of a three-terminal nonlinear resistor R and a point Q on the curves with coordinates $(I_{1_Q},V_{1_Q},I_{2_Q},V_{2_Q})$, we would like to find the most general three-terminal linear biasing network N so that when it is connected to R, as shown in Fig. 8-3*a*, the operating point will be at Q.

Observe that since the input- or the output-characteristic curves of any three-terminal nonlinear resistor can relate only three out of the four terminal variables, the operating point Q must be specified in both sets of curves. Moreover, since two of the three coordinates are identical in each set of curves, only two of the four coordinates of the operating point can be specified arbitrarily. This means that once we fix a point in either the input- or the output-characteristic curves, the corresponding point in the other set of characteristic curves must be chosen so that two of the coordinates of each point are consistent with one another. For example, consider the input- and output-characteristic curves of a typical unijunction transistor as shown in Fig. 8-3*b* and *c*. If we fix the operating point Q of the input curves at $I_1 = 6$ ma and $V_1 = 3$ volts, then Fig. 8-3*b* shows that this requires $V_2 = 15$ volts. Since the two coordinates V_2 and I_1 also appear in the output curves, it is clear that the corresponding point in Fig. 8-3*c* must be fixed at $V_2 = 15$ volts, $I_1 = 6$ ma, and $I_2 = 6.5$ ma. The point we are driving at is that we cannot arbitrarily specify any four numbers and require them to be the coordinates I_1, V_1, I_2, V_2 of the operating point Q because these coordinates may not be consistent with the characteristic curves of the three-terminal resistor. This is, of course, analogous to the case of a two-terminal resistor where we can only specify the current or the voltage; otherwise, the point may not be consistent with the v-i curve of the two-terminal resistor.

Since the biasing network N is assumed to be linear, its input- and output-characteristic curves consist of a family of parallel straight lines; namely,[1]

$$v_1 = r_{11}i_1 + r_{12}i_2 + e_1 \tag{8-3}$$

$$v_2 = r_{21}i_1 + r_{22}i_2 + e_2 \tag{8-4}$$

where r_{11}, r_{12}, r_{21}, r_{22}, e_1, and e_2 are constants which depend on the slopes and intercepts of the family of straight lines. If the biasing network contains only linear resistances and independent

---

[1] Review the resistance representation for a linear network given earlier in Eq. (3-17).

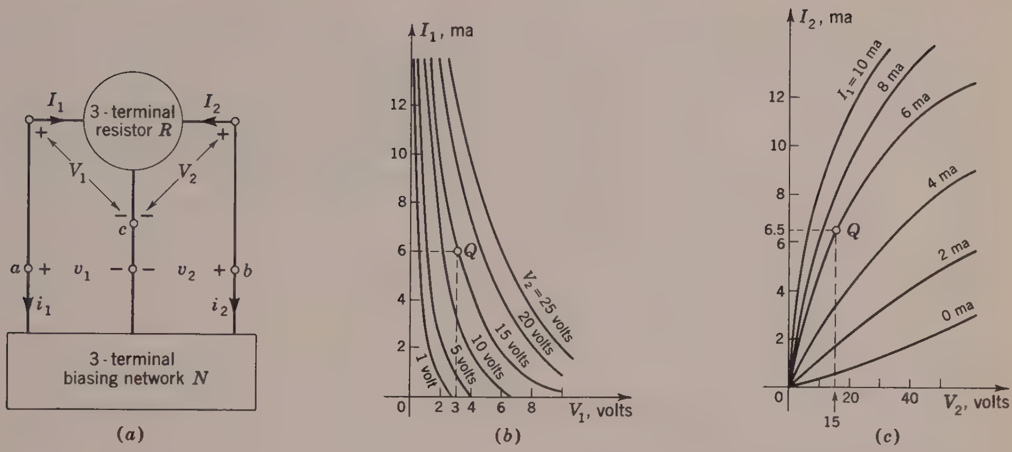


Fig. 8-3. A three-terminal biasing network is needed to establish an operating point Q on the input- and output-characteristic curves of a three-terminal resistor.

sources, it can be proved that $r_{12} = r_{21}$. Hence, without loss of generality, we can write Eqs. (8-3) and (8-4) in the form

$$v_1 = r_{11}i_1 + r_{12}i_2 + e_1 \tag{8-5}$$

$$v_2 = r_{12}i_1 + r_{22}i_2 + e_2 \tag{8-6}$$

Observe that Eqs. (8-5) and (8-6) can be written once we know r_{11}, r_{12}, r_{22}, e_1, and e_2. This means that the biasing network N is completely specified by these five parameters only. Our next objective is to determine these parameters so that the operating point of the three-terminal resistor of Fig. 8-3*a* will be located at $(I_{1_Q}, V_{1_Q}, I_{2_Q}, V_{2_Q})$. Since $i_1 = -I_1$, $i_2 = -I_2$, $v_1 = V_1$, and $v_2 = V_2$ at the operating point, Eqs. (8-5) and (8-6) become

$$V_{1_Q} = -r_{11}I_{1_Q} - r_{12}I_{2_Q} + e_1 \tag{8-7}$$

$$V_{2_Q} = -r_{12}I_{1_Q} - r_{22}I_{2_Q} + e_2 \tag{8-8}$$

Since there are five unknowns but only two equations, we can arbitrarily assume values for three of them and solve for the remaining two.[1]

What we have shown so far is that, given the coordinates $(I_{1_Q}, V_{1_Q}, I_{2_Q}, V_{2_Q})$ of any operating point Q, we can always find the five pertinent parameters r_{11}, r_{12}, r_{22}, e_1, and e_2 of the biasing network N. Our next objective is to find a general network configuration that can be designed with any prescribed set of parameters. It is clear that since there are five parameters to satisfy, any general network configuration must contain at least five elements. Our first instinct is to attempt to generalize the two-terminal biasing network of Fig. 8-1*b* to the three-terminal network shown in Fig. 8-4*a*. Observe, however, that we can always apply the v-shift theorem to eliminate the voltage source E_C by combining it with E_A and E_B. In other words, the five-element network shown in Fig. 8-4*b* is just as general as that shown in Fig. 8-4*a*. To find the values of the elements, we write the terminal equations of this network; thus

---

[1] In order for the equations to be consistent, the three assumed variables must not come from the same equation. Hence, the combinations (r_{11}, r_{12}, e_1) and (r_{12}, r_{22}, e_2) are not allowed.

$$v_1 = (R_1 + R_3)i_1 + R_3 i_2 + E_1 \tag{8-9}$$

$$v_2 = R_3 i_1 + (R_2 + R_3)i_2 + E_2 \tag{8-10}$$

Comparing Eqs. (8-9) and (8-10) with Eqs. (8-5) and (8-6), we obtain

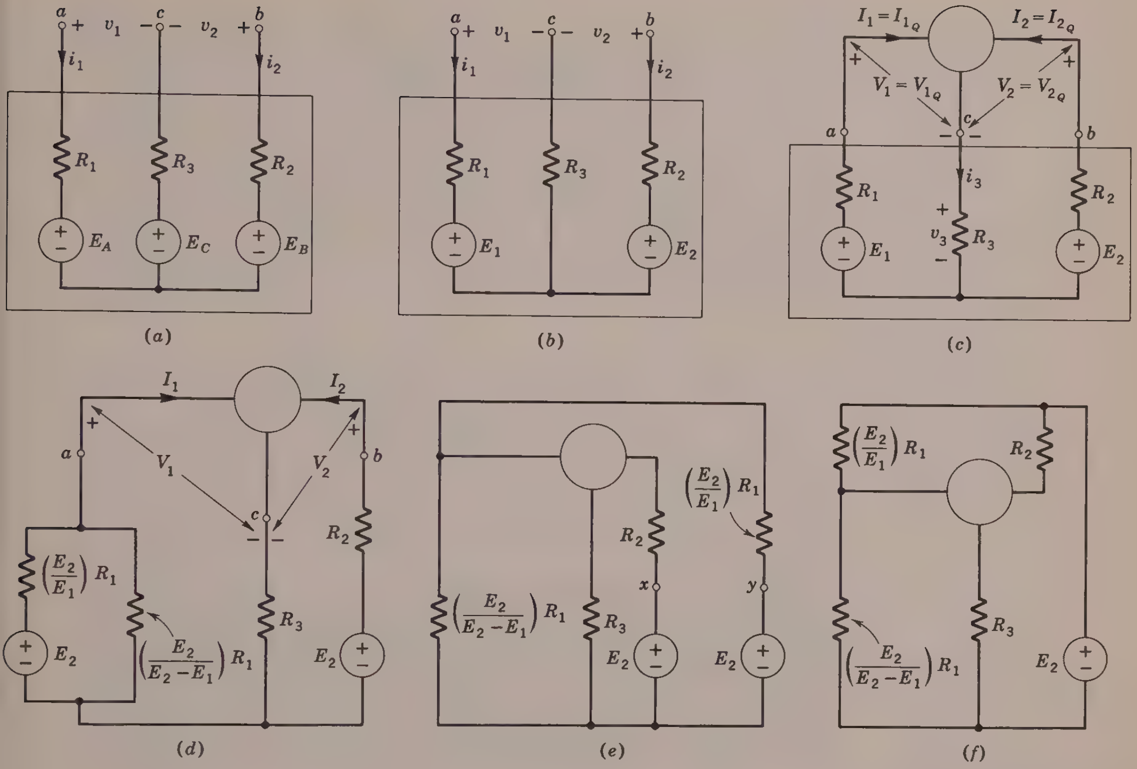
$$\begin{aligned} R_1 + R_3 &= r_{11} \\ R_3 &= r_{12} \\ R_2 + R_3 &= r_{22} \\ E_1 &= e_1 \\ E_2 &= e_2 \end{aligned} \tag{8-11}$$

From these equations we obtain

$$\begin{aligned} R_1 &= r_{11} - r_{12} \\ R_2 &= r_{22} - r_{12} \\ R_3 &= r_{12} \\ E_1 &= e_1 \\ E_2 &= e_2 \end{aligned} \tag{8-12}$$

We conclude, therefore, that we can always choose R_1, R_2, R_3, E_1, and E_2 to realize any specification r_{11}, r_{12}, r_{22}, e_1, and e_2. If we connect the network of Fig. 8-4*b* to the three-terminal resistor shown in Fig. 8-4*c*, we obtain the prescribed operating point Q.

A closer examination of Eqs. (8-11) and (8-12) shows that there is a one-to-one correspondence between R_1, R_2, R_3, E_1, E_2, and r_{11}, r_{12}, r_{22}, e_1, e_2. Given either set of parameters, we can compute the other. This means that if we can arbitrarily assume the values of three of the five parameters r_{11}, r_{12}, r_{22}, e_1, e_2, we can also arbitrarily assume the values of three of the five elements R_1, R_2, R_3, E_1, E_2. The inconsistent combination (r_{11}, r_{12}, e_1) corre-



(a) (b) (c) (d) (e) (f)

Fig. 8-4. The derivation of general biasing networks for three-terminal resistors.

sponds to (R_1,R_3,E_1) and must be avoided. Similarly, the inconsistent combination (r_{22},r_{12},e_2) corresponds to (R_2,R_3,E_2) and must be avoided. Any other combination is acceptable. For example, if we assume arbitrary values for the combination (R_1,E_1,E_2), we can compute the remaining two parameters R_2 and R_3 as follows: Substituting Eq. (8-11) into Eqs. (8-7) and (8-8), we obtain

$$V_{1_Q} = -(R_1 + R_3)I_{1_Q} - R_3 I_{2_Q} + E_1 \tag{8-13}$$

$$V_{2_Q} = -R_3 I_{1_Q} - (R_2 + R_3)I_{2_Q} + E_2 \tag{8-14}$$

Solving for the unknowns R_3 and R_2, we obtain

$$R_3 = \frac{E_1 - V_{1_Q} - R_1 I_{1_Q}}{(I_{1_Q} + I_{2_Q})} \tag{8-15}$$

$$R_2 = \frac{E_2 - V_{2_Q} - R_3(I_{1_Q} + I_{2_Q})}{I_{2_Q}} \tag{8-16}$$

Observe that we could also obtain these two relations directly from the network of Fig. 8-4*c*. For example, the voltage across R_3 is obtained by applying KVL to obtain

$$v_3 = E_1 - V_{1_Q} - R_1 I_{1_Q} \tag{8-17}$$

Similarly, the current through R_3 is obtained by applying the generalized KCL to obtain

$$i_3 = (I_{1_Q} + I_{2_Q}) \tag{8-18}$$

The value of R_3 is, therefore, obtained by dividing Eq. (8-17) by Eq. (8-18). The value for R_2 can be similarly obtained.

If we assume another combination of three unknowns, we can determine the remaining two unknowns by either of these two methods. For future reference, the relationships corresponding to the 10 possible combinations are tabulated in column 1 of Table 8-1.[1]

Although the biasing network of Fig. 8-4*c* is completely general, it has the disadvantage that two voltage sources are generally required. We shall now show that by applying the principle of equivalent networks, it is possible to trade one of the two voltage sources for a linear resistor. To do this, let us replace the series combination of R_1 and E_1 of Fig. 8-4*c* by the equivalent network shown in Fig. 8-4*d*. This can be redrawn into the form shown in Fig. 8-4*e*. Now observe that the voltage between terminals *x-y* is zero, and hence we can short-circuit terminals *x-y*, thus combining the two identical voltage sources into one as shown in Fig. 8-4*f*. We have shown that this network is equivalent to the network of Fig. 8-4*c* and we shall, for convenience, rename the resistance $[E_2/(E_2 - E_1)]R_1$ by R_1, the resistance $(E_2/E_1)R_1$ by R_4, and the voltage source E_2 by E. It is clear that we can assume arbitrary values for three of the five elements R_1, R_2, R_3, R_4, and E and solve for the remaining two unknowns. The relations corresponding to the 10 possible combinations are tabulated in column 2 of Table 8-1.

Another common equivalent circuit of Fig. 8-4*c* is shown in column 3 of Table 8-1. Again, we can assume arbitrary values for three of the five elements and compute the remaining two elements. The relations corresponding to the 10 possible combinations are tabulated in column 3 of Table 8-1. This table is extremely useful because it is applicable to any three-terminal resistor. Let us illustrate the use of this table by some numerical examples.

---

[1] The reader is encouraged to derive the entries in this table *directly* from the network shown in Fig. 8-4*c* in a manner similar to the derivation of Eqs. (8-17) and (8-18). The ability to do this quickly will make it unnecessary to depend on the table.

EXAMPLES

1. Suppose we want to bias the unijunction transistor curves in Fig. 8-3 at the operating point Q with coordinates $I_{1_Q} = 6$ ma, $V_{1_Q} = 3$ volts, $I_{2_Q} = 6.5$ ma, and $V_{2_Q} = 15$ volts.

   Suppose we choose the biasing network corresponding to column 1 of Table 8-1 and arbitrarily choose $E_1 = 9$ volts, $E_2 = 28$ volts, and $R_3 = 0\ \Omega$. This combination corresponds to row 2 of column 1. Therefore, we can immediately determine R_1 and R_2 as follows:

$$R_1 = \frac{9 - 3 - 0(12.5)(10^{-3})}{6(10^{-3})} = 1\text{ k}\Omega$$

$$R_2 = \frac{28 - 15 - 0(12.5)(10^{-3})}{6.5(10^{-3})} = 2\text{ k}\Omega$$

   The resulting biasing network is shown in Fig. 8-5*a*.

2. Let $I_B = 0.45$ ma, $V_{BE} = 0.25$ volt, $I_C = 25$ ma, and $V_{CE} = 6.25$ volts be the coordinates of a point Q on the common-emitter characteristic curves of an *n-p-n* transistor. To bias the transistor at Q, let us arbitrarily choose the biasing network shown in column 2 of Table 8-1. Suppose that only a 15-volt battery is available, and that from other practical considerations it is required that $R_2 = 100\ \Omega$ and $R_4 = 10\text{ k}\Omega$. This combination corresponds to row 6 of column 2. Therefore, we immediately obtain

$$R_1 = \frac{10^4[15-(6.25-0.25)-(100)(25)(10^{-3})]}{6.25-0.25+(100)(25)(10^{-3})-(10^4)(0.45)(10^{-3})} = 16.25\text{ k}\Omega$$

$$R_3 = \frac{15-(100)(25)(10^{-3})-6.25}{(0.45+25)(10^{-3})} \approx 246\ \Omega$$

   The resulting biasing network is shown in Fig. 8-5*b*.

3. Instead of an *n-p-n* transistor, let us choose a *p-n-p* transistor whose operating point differs from the above example by only a negative sign; namely, $I_B = -0.45$ ma, $V_{BE} = -0.25$ volt, $I_C = -25$ ma, and $V_{CE} = -6.25$ volts. In view of the complementary-network theorem, we immediately obtain the biasing circuit shown in Fig. 8-5*c*.

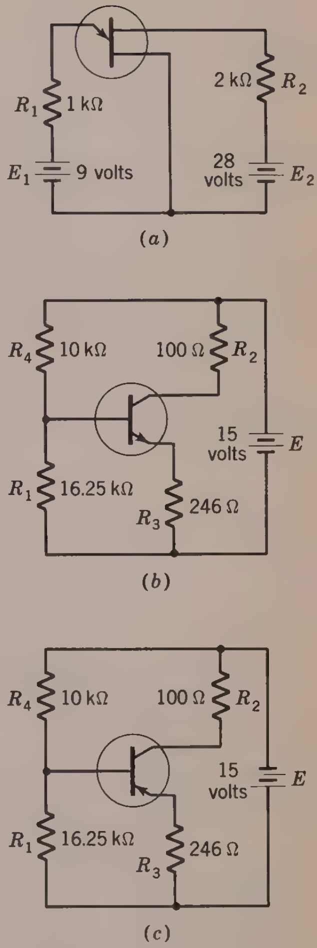
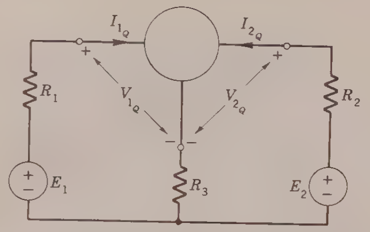
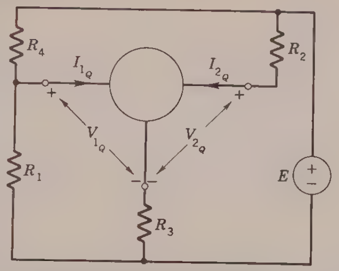
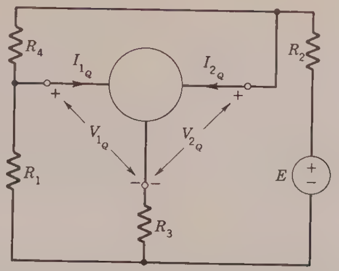


Fig. 8-5. The complete biasing circuits for a typical unijunction, *n-p-n* transistor, and *p-n-p* transistor.

In theory, the three biasing networks given in Table 8-1 are completely equivalent to each other. That is to say, it is possible to choose appropriate values for the elements in each circuit such

TABLE 8-1 *Design equations for three common biasing circuits.*

	 I_{1Q} I_{2Q} + + R_1 V_{1Q} V_{2Q} R_2 − − E_1 R_3 E_2	 R_4 R_2 I_{1Q} I_{2Q} + + V_{1Q} V_{2Q} R_1 E − − R_3	 R_4 R_2 I_{1Q} I_{2Q} + + V_{1Q} V_{2Q} R_1 E − − R_3
	Biasing-circuit configuration 1	Biasing-circuit configuration 2	Biasing-circuit configuration 3
1	Given: E_1, E_2, R_1 $R_3 = \frac{E_1 - V_{1Q} - R_1 I_{1Q}}{(I_{1Q} + I_{2Q})}$ $R_2 = \frac{E_2 - V_{2Q} - R_3(I_{1Q} + I_{2Q})}{I_{2Q}}$	Given: E, R_1, R_3 $R_2 = \frac{E - V_{2Q} - R_3(I_{1Q} + I_{2Q})}{I_{2Q}}$ $R_4 = \frac{[E - V_{1Q} - (I_{1Q} + I_{2Q})R_3]R_1}{V_{1Q} + (I_{1Q} + I_{2Q})R_3 + R_1 I_{1Q}}$	Given: E, R_1, R_3 $R_2 = \frac{[E - V_{2Q} - (I_{1Q} + I_{2Q})R_3]R_1}{V_{1Q} + (R_1 + R_3)(I_{1Q} + I_{2Q})}$ $R_4 = \frac{R_1(V_{2Q} - V_{1Q})}{V_{1Q} + R_3(I_{1Q} + I_{2Q}) + R_1 I_{1Q}}$
2	Given: E_1, E_2, R_3 $R_1 = \frac{E_1 - V_{1Q} - R_3(I_{1Q} + I_{2Q})}{I_{1Q}}$ $R_2 = \frac{E_2 - V_{2Q} - R_3(I_{1Q} + I_{2Q})}{I_{2Q}}$	Given: E, R_1, R_2 $R_3 = \frac{E - V_{2Q} - R_2 I_{2Q}}{(I_{1Q} + I_{2Q})}$ $R_4 = \frac{(V_{2Q} - V_{1Q} + R_2 I_{2Q})R_1}{V_{1Q} - V_{2Q} + E - R_2 I_{2Q} + R_1 I_{1Q}}$	Given: E, R_1, R_2 $R_3 = \frac{R_1[E - V_{2Q} - R_2(I_{1Q} + I_{2Q})]R_2 V_{1Q}}{(R_1 + R_2)(I_{1Q} + I_{2Q})}$ $R_4 = \frac{(R_1 + R_2)(V_{2Q} - V_{1Q})}{E + (V_{1Q} - V_{2Q}) + R_1 I_{1Q} - R_2 I_{2Q}}$
3	Given: E_1, E_2, R_2 $R_1 = \frac{E_1 - V_{1Q} - E_2 + V_{2Q} + R_2 I_{2Q}}{I_{1Q}}$ $R_3 = \frac{E_2 - V_{2Q} - R_2 I_{2Q}}{(I_{1Q} + I_{2Q})}$	Given: E, R_1, R_4 $R_3 = \frac{(E - V_{1Q} - R_4 I_{1Q})R_1 - R_4 V_{1Q}}{(R_1 + R_4)(I_{1Q} + I_{2Q})}$ $R_2 = \frac{(R_1 + R_4)(V_{1Q} - V_{2Q}) + R_4(E + R_1 I_{1Q})}{(R_1 + R_4)I_{2Q}}$	Given: E, R_1, R_4 $R_3 = \frac{R_1(V_{2Q} - V_{1Q}) - R_4(R_1 I_{1Q} + V_{1Q})}{R_4(I_{1Q} + I_{2Q})}$ $R_2 = \frac{(V_{1Q} - V_{2Q})(R_1 + R_4) + R_4(E + R_1 I_{1Q})}{(V_{2Q} - V_{1Q}) + R_4 I_{2Q}}$
4	Given: E_1, R_1, R_3 Inconsistent	Given: E, R_2, R_3 Inconsistent	Given: E, R_2, R_3 $R_1 = \frac{R_2[V_{1Q} + R_3(I_{1Q} + I_{2Q})]}{E - V_{2Q} - (R_2 + R_3)(I_{1Q} + I_{2Q})}$ $R_4 = \frac{R_2(V_{2Q} - V_{1Q})}{E - V_{2Q} - R_3(I_{1Q} + I_{2Q}) - R_2 I_{2Q}}$

5	Given: E_1, R_1, R_2 $R_3 = \dfrac{E_1 - V_{1Q} - R_1 I_{1Q}}{(I_{1Q} + I_{2Q})}$ $E_2 = E_1 + (V_{2Q} - V_{1Q}) + R_2 I_{2Q} - R_1 I_{1Q}$	Given: E, R_3, R_4 $R_2 = \dfrac{E - V_{2Q} - R_3(I_{1Q} + I_{2Q})}{I_{2Q}}$ $R_1 = \dfrac{R_4[V_{1Q} + R_3(I_{1Q} + I_{2Q})]}{E - V_{1Q} - R_3(I_{1Q} + I_{2Q}) - R_4 I_{1Q}}$	Given: E, R_3, R_4 $R_1 = \dfrac{R_4[V_{1Q} + (I_{1Q} + I_{2Q})R_3]}{V_{2Q} - V_{1Q} - R_4 I_{1Q}}$ $R_2 = \dfrac{R_4[E - V_{2Q} - R_3(I_{1Q} + I_{2Q})]}{V_{2Q} - V_{1Q} + R_4 I_{2Q}}$
6	Given: E_1, R_2, R_3 $R_1 = \dfrac{E_1 - V_{1Q} - R_3(I_{1Q} + I_{2Q})}{I_{1Q}}$ $E_2 = V_{2Q} + R_3(I_{1Q} + I_{2Q}) + R_2 I_{2Q}$	Given: E, R_2, R_4 $R_1 = \dfrac{R_4[E - (V_{2Q} - V_{1Q}) - R_2 I_{2Q}]}{V_{2Q} - V_{1Q} + R_2 I_{2Q} - R_4 I_{1Q}}$ $R_3 = \dfrac{E - R_2 I_{2Q} - V_{2Q}}{(I_{1Q} + I_{2Q})}$	Given: E, R_2, R_4 $R_1 = \dfrac{(R_2 + R_4)(V_{1Q} - V_{2Q}) + R_4(E - R_2 I_{2Q})}{(V_{2Q} - V_{1Q}) - R_4 I_{1Q}}$ $R_3 = \dfrac{R_4(E - V_{2Q}) - R_2(V_{2Q} - V_{1Q} + R_4 I_{2Q})}{R_4(I_{1Q} + I_{2Q})}$
7	Given: E_2, R_1, R_3 $R_2 = \dfrac{E_2 - V_{2Q} - R_3(I_{1Q} + I_{2Q})}{I_{2Q}}$ $E_1 = V_{1Q} + R_1 I_{1Q} + R_3(I_{1Q} + I_{2Q})$	Given: R_1, R_2, R_3 $R_4 = \dfrac{R_1[R_2 I_{2Q} + (V_{2Q} - V_{1Q})]}{V_{1Q} + (I_{1Q} + I_{2Q})R_3 + R_1 I_{1Q}}$ $E = V_{2Q} + R_2 I_{2Q} + R_3(I_{1Q} + I_{2Q})$	Given: R_1, R_2, R_3 $R_4 = \dfrac{R_1(V_{2Q} - V_{1Q})}{R_1 I_{1Q} + V_{1Q} + R_3(I_{1Q} + I_{2Q})}$ $E = (I_{1Q} + I_{2Q})R_3 + V_{2Q} + R_2\left(I_{2Q} + \dfrac{V_{2Q} - V_{1Q}}{R_4}\right)$
8	Given: E_2, R_1, R_2 $R_3 = \dfrac{E_2 - V_2 - R_2 I_{2Q}}{(I_{1Q} + I_{2Q})}$ $E_1 = E_2 + V_{1Q} - V_{2Q} + R_1 I_{1Q} - R_2 I_{2Q}$	Given: R_1, R_3, R_4 $R_2 = \dfrac{R_1(V_{1Q} - V_{2Q}) + R_4[V_{1Q} + R_1 I_{1Q} + R_3(I_{1Q} + I_{2Q})]}{R_1 I_{2Q}}$ $E = [R_3(I_{1Q} + I_{2Q}) + V_{1Q}]\left(1 + \dfrac{R_4}{R_1}\right) + R_4 I_{1Q}$	Given: R_1, R_3, R_4 Inconsistent
9	Given: E_2, R_2, R_3 Inconsistent	Given: R_1, R_2, R_4 $R_3 = \dfrac{R_1(V_{2Q} - V_{1Q} + R_2 I_{2Q} - R_4 I_{1Q}) - R_4 V_{1Q}}{R_4(I_{1Q} + I_{2Q})}$ $E = (V_{2Q} - V_{1Q} + R_2 I_{2Q})\left(1 + \dfrac{R_1}{R_4}\right) - R_1 I_{1Q}$	Given: R_1, R_2, R_4 $R_3 = \dfrac{[(V_{2Q} - V_{1Q}) - R_4 I_{1Q}]R_1 - R_4 V_{1Q}}{R_4(I_{1Q} + I_{2Q})}$ $E = \dfrac{(V_{2Q} - V_{1Q})(R_1 + R_2 + R_4)}{R_4} - R_1 I_{1Q} + R_2 I_{2Q}$
10	Given: R_1, R_2, R_3 $E_1 = V_{1Q} + R_1 I_{1Q} + R_3(I_{1Q} + I_{2Q})$ $E_2 = V_{2Q} + R_2 I_{2Q} + R_3(I_{1Q} + I_{2Q})$	Given: R_2, R_3, R_4 $R_1 = \dfrac{R_4[V_{1Q} + R_3(I_{1Q} + I_{2Q})]}{(V_{2Q} - V_{1Q} + R_2 I_{2Q} - R_4 I_{1Q})}$ $E = R_3(I_{1Q} + I_{2Q}) + V_{2Q} + R_2 I_{2Q}$	Given: R_2, R_3, R_4 $R_1 = \dfrac{R_4[R_3(I_{1Q} + I_{2Q}) + V_{1Q}]}{V_{2Q} - V_{1Q} - R_4 I_{1Q}}$ $E = \dfrac{(V_{2Q} - V_{1Q})R_2}{R_4} + R_3(I_{1Q} + I_{2Q}) + R_2 I_{2Q} + V_{2Q}$

that, externally, we could not tell the difference. Why, then, do we need three circuits? There are many practical constraints which make this necessary. For example, sometimes it may be necessary to use only one battery, in which case the circuit shown in column 2 or column 3 may be chosen. The choice of the circuit in column 2 or column 3 may further depend on other practical considerations. For example, one circuit might require a higher-voltage battery than the other, or another circuit might be draining more power from the battery than the other. Moreover, the formulas given in Table 8-1 guarantee only a solution, but not the most practical solution. Indeed, it is possible to end with negative resistances in some cases! Whenever such impractical values are obtained it is obviously desirable to either pick a different configuration or change some of the prescribed parameters.

8-2-3 SENSITIVITY CONSIDERATION IN THREE-TERMINAL BIASING CIRCUITS

We have shown that a three-terminal biasing network generally needs five elements, the values of three of which can be arbitrarily chosen (except the two inconsistent combinations). In practice, we may take advantage of these extra degrees of freedom to optimize the network with respect to some engineering criterion.

One of the most important criterions is to minimize the sensitivity of the operating point with respect to variations in temperature. We have already been introduced to the thermal-runaway phenomenon. This phenomenon is more troublesome in three-terminal nonlinear resistors and must always be avoided. In order to study this phenomenon, let us rewrite Eqs. (8-5) and (8-6) of the biasing network in terms of the terminal variables of the three-terminal resistor

$$V_1 = -r_{11}I_1 - r_{12}I_2 + e_1 \tag{8-19}$$

$$V_2 = -r_{12}I_1 - r_{22}I_2 + e_2 \tag{8-20}$$

Remember that for each biasing network, the parameters r_{11}, r_{12}, r_{22}, e_1, and e_2 are fixed numbers. We can represent Eq. (8-19) as a family of parallel load lines in the V_1-I_1 plane as shown in Fig. 8-6*a*, where I_2 is chosen as the parameter. Similarly, we can represent Eq. (8-20) as a family of parallel load lines in the V_2-I_2 plane as shown in Fig. 8-6*b*, where I_1 is chosen as the parameter. Observe that when $r_{12} = 0$, the two families of load lines reduce to the two load lines corresponding to $I_2 = 0$ in Fig. 8-6*a* and

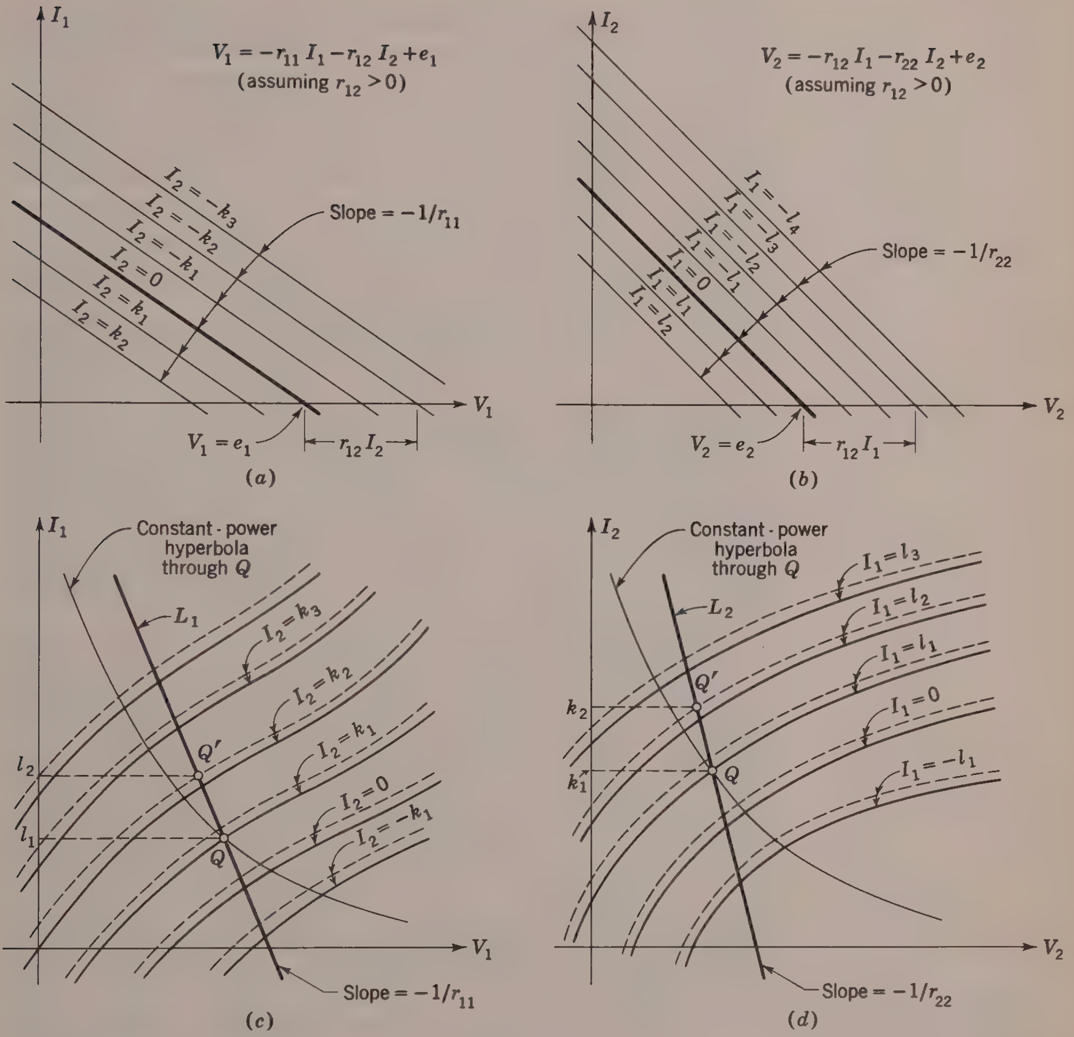


Fig. 8-6. A graphical illustration of the two interacting mechanisms responsible for the thermal-runaway phenomenon commonly observed in a three-terminal resistor.

$I_1 = 0$ in Fig. 8-6*b*. From Eq. (8-12) we observe that $r_{12} = 0$ implies $R_3 = 0$ in the biasing network of Fig. 8-4*b*.

Let us now consider how thermal runaway can occur when $r_{12} = 0$. Consider, for example, the two families of negative-temperature-coefficient characteristic curves shown in Figs. 8-6*c* and 8-6*d*. The solid curves represent the characteristics at temperature T_1, while the dotted curves represent the characteristics at temperature T_2, where $T_2 > T_1$.

Consider first the load lines L_1 in Fig. 8-6*c* and L_2 in Fig. 8-6*d*. At $T = T_1$, the operating point will be at Q in Fig. 8-6*c*, where

$I_2 = k_1$. Similarly, the operating point will be at Q in Fig. 8-6*d*, where $I_1 = l_1$. When the temperature increases to T_2, the curves corresponding to $I_2 = k_1$ and $I_1 = l_1$ shift upward as shown by the dotted curves. This will result in a relocation of the operating point. Let us suppose that the new operating point Q' in Fig. 8-6*c* is at the intersection between the load line L_1 and the new dotted curve $I_2 = k_2$ (at T_2). Similarly, let us suppose the new operating point Q' in Fig. 8-6*d* is at the intersection between the load line L_2 and the new dotted curve $I_1 = l_2$ (at T_2). At this point, two interacting phenomena may occur simultaneously.

First, if Q' lies above the constant-power hyperbola through Q in Fig. 8-6*c* and 8-6*d*, then the power dissipation at temperature T_2 will be greater than the dissipation at T_1. This may induce a self-heating effect and further raise the temperature, thus moving the operating point even further up the load line. This will further increase the heating effect, and the result is a cumulative rise in temperature. When the operating point moves beyond the maximum-power hyperbola, the device will be destroyed. (The above mechanism is similar to that encountered earlier in the biasing of two-terminal resistors.)

In addition to the cumulative-heating mechanism, there is a second mechanism which hastens the thermal-runaway phenomenon. To see this, let us examine the operating points Q and Q' again. Observe that the value of the parameter I_2 (Fig. 8-6*c*) at Q' is not the same as the value at Q. In particular, we found $I_2(T = T_2) > I_2(T = T_1)$. Similarly, the value of the parameter I_1 (Fig. 8-6*d*) at Q' is not the same as the value at Q. In particular, $I_1(T = T_2) > I_1(T = T_1)$. The effect of an increase in the value of the parameters I_2 and I_1 is that the new operating point will move up higher than if the parameters remain constant. This introduces a push-pull effect which hastens the thermal-runaway phenomenon.

Of course, the above two mechanisms are not really separable because they occur simultaneously. However, we can avoid thermal runaway by treating these two mechanisms separately.

Since the first thermal-runaway mechanism in the three-terminal biasing circuits is identical with that in the two-terminal biasing circuits, we can generalize the earlier criterion as follows:

FIRST THERMAL-RUNAWAY CRITERION FOR THREE-TERMINAL RESISTORS

Let $(I_{1_Q}, V_{1_Q}, I_{2_Q}, V_{2_Q})$ be the prescribed operating point of a three-terminal nonlinear resistor. In order to prevent the cumulative-heating mechanism from occurring in a biasing circuit where $r_{12} = 0$, it is sufficient that

$$r_{11} > \frac{V_{1Q}}{I_{1Q}} \tag{8-21}$$

$$r_{22} > \frac{V_{2Q}}{I_{2Q}} \tag{8-22}$$

if the resistor has a negative temperature coefficient. Conversely, if the resistor has a positive temperature coefficient, reverse the inequality signs in Eqs. (8-21) and (8-22).

Observe that the above criterion is stated for the case $r_{12} = 0$. If $r_{12} \neq 0$, the voltage intercepts of the load lines L_1 and L_2 will depend on the value of the parameters I_2 and I_1 as shown in Fig. 8-6*a* and *b*. However, since the slope of these load lines remains the same, the criterion will still be valuable in determining the appropriate conditions.

In order to prevent or minimize the second thermal-runaway mechanism, it is necessary to move the load lines in a direction to minimize the change in the values of I_1 and I_2. For example, in the characteristic curves shown in Fig. 8-6*c* it is desirable that as the value of the parameter I_2 increases, the load line L_1 moves to the left. This will result in a smaller increase in I_1. Similarly, in the characteristic curves shown in Fig. 8-6*d* it is desirable that as the value of the parameter I_1 increases, the load line L_2 moves to the left. This will result in a smaller increase in I_2. From Fig. 8-6*a* and *b* we observe that in order to move L_1 to the left, it is necessary and sufficient that $r_{12} > 0$. Similarly, in order to move L_2 to the left, it is necessary and sufficient that $r_{12} > 0$. In either case, we conclude that $r_{12} \neq 0$. This observation is important enough to be stated as a theorem:[1]

SECOND THERMAL-RUNAWAY CRITERION FOR THREE-TERMINAL RESISTORS

In order to minimize the effect of the second thermal-runaway mechanism, it is necessary that $r_{12} \neq 0$. In particular, if x_1 is the parameter of the characteristic curve in the V_1-I_1 plane, and x_2 is the parameter of the characteristic curve in the V_2-I_2 plane, then r_{12} must be chosen such that

1. The load line L_1 will move in such a direction as to minimize the change in the value of x_2 in the V_1-I_1 plane.

---

[1] In the early days of transistors the biasing network shown in Fig. 8-4*b* with $R_3 = r_{12} = 0$ was often chosen because this circuit had been used successfully for biasing vacuum tubes. It took a while, after burning out many transistors, for engineers to realize that a transistor has a different temperature characteristic, and that the resistor R_3 should always be inserted to prevent a thermal runaway.

2. The load line L_2 will move in such a direction as to minimize the change in the value of x_1 in the V_2-I_2 plane.

The first and second thermal-runaway criteria for three-terminal resistors are valuable in the design of practical biasing circuits. For example, if we want to bias a temperature-sensitive nonlinear resistor such as a transistor by the biasing circuit shown in column 1 of Table 8-1, it is important that we do not make $R_3 = 0$. This is so because $R_3 = r_{12}$ for this circuit. Given a specific three-terminal resistor, it is possible to derive a number of shortcuts and rules of thumb. However, since no new concept is involved, we shall leave the details of this to a more design-oriented course.

Exercise 1: When the solar cell is used as a power source, the external biasing network is usually just a load resistor. Since the power available is very small, it is important that we bias the solar cell at an optimal point Q in order to obtain maximum power dissipation in the load resistor. (*a*) Find a systematic graphical procedure for locating the optimum point corresponding to each level of light illumination. (*b*) Specify the DP plot of the *passive* load resistor so that maximum power dissipation is attained for all levels of light illumination for the solar cell type 1NXX11.

Exercise 2: Discuss the thermal-runaway phenomenon in a positive temperature coefficient two-terminal resistor.

Exercise 3: Derive the biasing equations listed in column 1 of Table 8-1 directly from the circuit by writing KVL and KCL equations.

Exercise 4: Repeat Exercise 3 for column 2 of Table 8-1.

Exercise 5: Repeat Exercise 3 for column 3 of Table 8-1.

Exercise 6: Find an example of a biasing network for which one or more resistors as determined from Table 8-1 may come out negative.

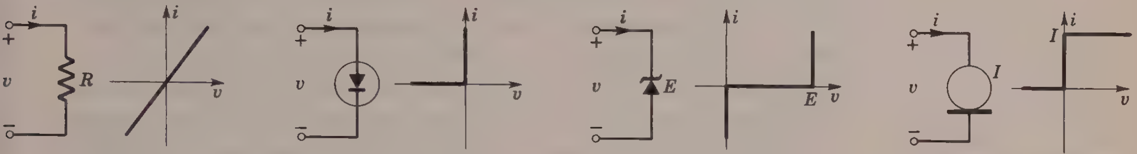
Exercise 7: Discuss the two mechanisms leading to thermal-runaway phenomena in a positive temperature coefficient three-terminal resistor.

8-3 SYNTHESIS OF DP PLOTS

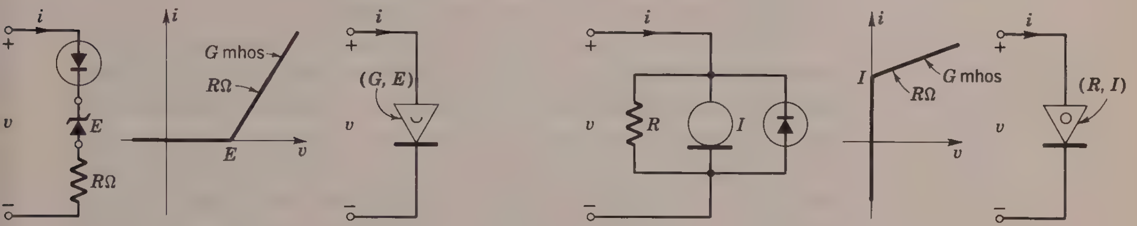
There are many occasions when it is necessary to synthesize a network having a prescribed DP plot. We have already seen many examples which demonstrate the usefulness of DP plots not only as a powerful analytical technique, but also as a basic approach to synthesis. In fact, much of the material in the remaining chapters of this book will depend heavily upon our ability to synthesize a network with a prescribed DP plot. It is important from a practical point of view that the "ingredients" or "building blocks" be commercially available or that they can be readily assembled using only commercially available components. We shall find

TABLE 8-2 *Basic building blocks for synthesis.*

PASSIVE COMPONENTS

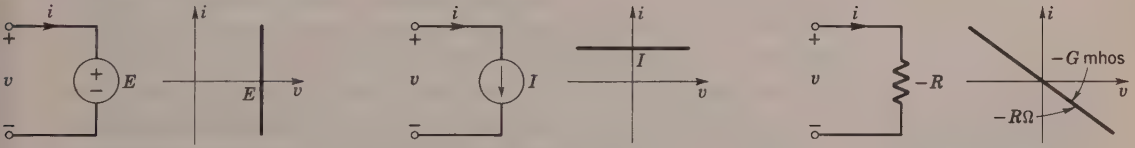


(*a*) Positive resistor (*b*) Junction diode (*c*) Zener diode (*d*) Constant-current diode

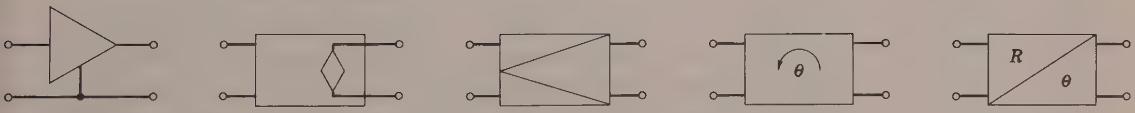


(*e*) Concave resistor (*f*) Convex resistor

ACTIVE COMPONENTS



(*a*) Independent voltage source (*b*) Independent current source (*c*) Negative resistor



(*d*) Operational amplifier (*e*) Controlled source (*f*) Scalor (*g*) Rotator (*h*) Reflector

it convenient to classify the building blocks into two groups, namely, *passive building blocks* and *active building blocks*.[1] A list of passive components that we shall use in this chapter is shown in the upper half of Table 8-2 with their somewhat idealized v-i curves.[2] The first four resistors, positive resistor, junction diode, zener diode, and constant-current diode, are commercially available components. By combining these basic components as shown in (*e*) and (*f*) of this table, we obtain the *concave resistor*

[1] Recall that a two-terminal resistor is passive if, and only if, its v-i curve lies only in the first and the third quadrants. Otherwise, it is active.

[2] For simplicity, we shall represent all v-i curves in piecewise-linear form. The discrepancy arising from this approximation can be easily taken care of after the synthesis is done.

and *convex resistor* defined earlier in Figs. 6-20 and 6-21. Since these two elements will be used most often, we shall denote them by the symbols shown on the right of the respective v-i curves. Notice that only two parameters are needed to completely specify these resistors, namely, the slope G and the voltage intercept E for the concave resistor and the reciprocal of the slope R and the current intercept I for the convex resistor.[1] Hence, it is unnecessary to draw the v-i curves of these elements. Notice that the junction diode (approximated by an ideal diode) can be considered as a special case of a concave resistor with $(G,E) = (\infty,0)$.

A list of active components is shown in the lower part of Table 8-2. The independent voltage source, independent current source, and the negative resistor are two-terminal active elements, whereas the operational amplifier, the controlled source (four possible types), the scalor, the rotator, and the reflector are four-terminal or two-port active elements. Among these components, the independent sources and the operational amplifiers are commercially available. The rest of the components may be easily synthesized in terms of appropriate transistor circuits and operational amplifiers (see Sec. 3-7).

From the practical point of view, passive components are preferred over active components not only because they are cheaper, but also because all active components require an external power supply, a nuisance, to say the least. Moreover, active components are usually more sensitive to element parameter variations. They also often lead to instability problems. These disadvantages do not, however, imply that these elements are not practical. On the contrary, most of the important practical circuits are active circuits, and, therefore, require the use of active elements. As a matter of fact, the advent of integrated circuits is quickly changing the designer's attitude because from the user's point of view, a seemingly complicated circuit can be integrated so that it is no larger than a passive component. We shall now discuss how the building blocks in Table 8-2 can be systematically combined into a network with a prescribed DP plot. We shall start with the simplest monotonic DP plot and then go to current-controlled and voltage-controlled DP plots.[2]

8-3-1 SYNTHESIS OF MONOTONIC DP PLOTS

A DP plot Γ is said to be monotonic if the slope at each point on Γ is nonnegative. A monotonic DP plot is said to be concave (convex) if the slope is a monotonically increasing (decreasing) function of the terminal voltage. Any monotonic DP plot may be broken up into a combination of concave and convex curves. This suggests that we should be able to synthesize any monotonic DP plot with concave resistors, convex resistors, and batteries.

---

[1] To avoid the clumsy notation $1/R$, we shall let the parameter associated with each segment of a piecewise-linear v-i curve denote either the slope G (conductance in mhos) or the reciprocal slope R (resistance in ohms) of the segment. We shall resolve any possible ambiguity by using the symbols G_j and g_j for conductances and the symbols R_j and r_j for resistances. When specific numbers are involved, the *units* in mhos or ohms will be attached.

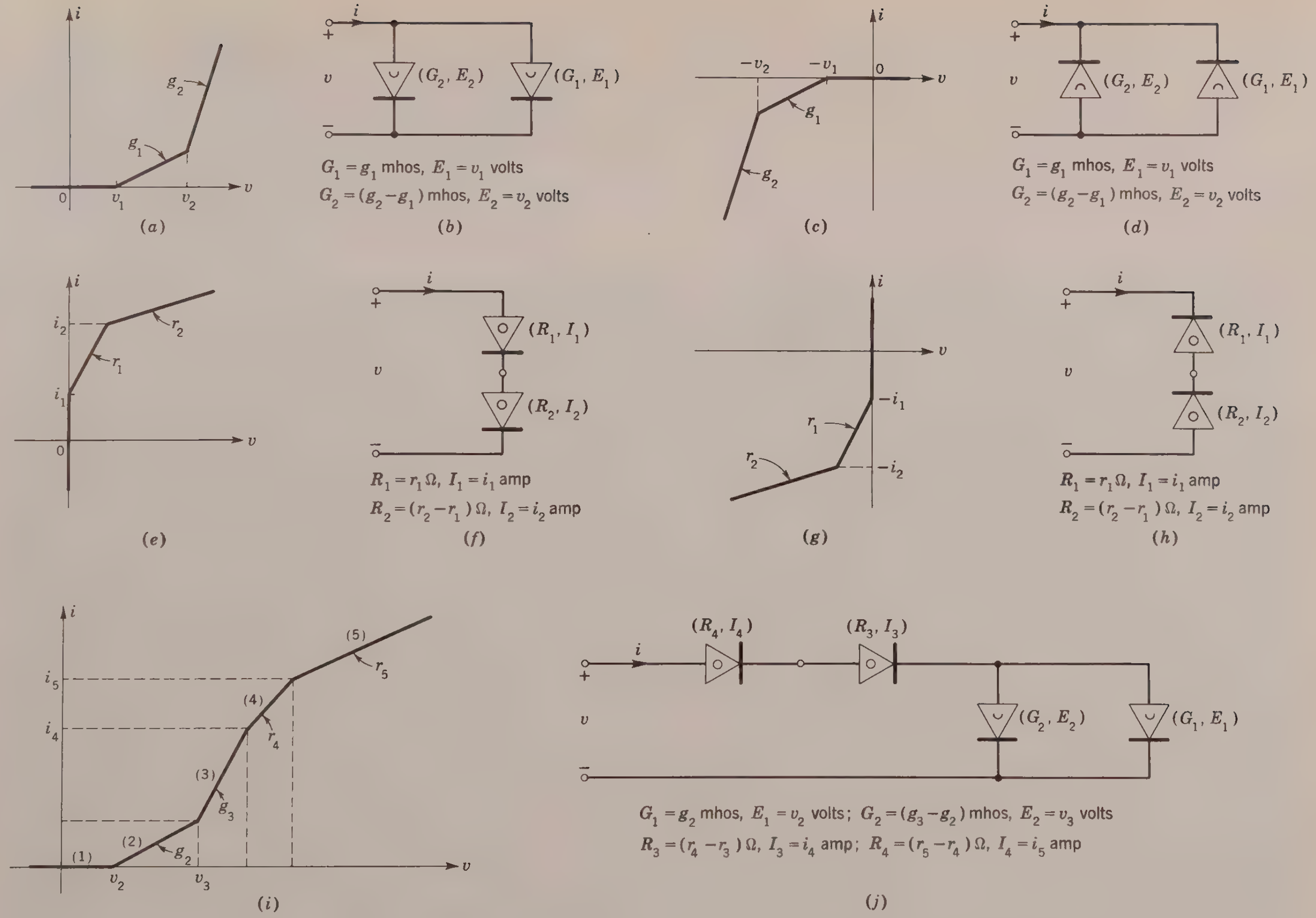
[2] The same principle can be extended to the synthesis of multivalued DP plots; however, we shall not discuss this more esoteric case in this book.

Consider first the concave DP plot shown in Fig. 8-7*a*. This can be realized by two concave resistors in parallel as shown in Fig. 8-7*b*. It is clear from the parallel-combination technique that the second concave resistor does not come into play until $v \geq v_2$. This means that if we connect a third concave resistor with $E = v_3 > v_2$, then the DP plot will remain the same as in Fig. 8-7*a* until $v \geq v_3$. In other words, each additional concave resistor introduces a new breakpoint and changes the slope of only the last segment beyond this breakpoint. It is a simple matter, therefore, to synthesize a concave DP plot with any number of segments in the first quadrant. Since any DP plot in the third quadrant can be obtained by rotating its complementary DP plot in the first quadrant by 180° with respect to the origin (Fig. 8-7*c*), it can be realized by the complementary network in accordance with the complementary-network theorem; namely, simply transpose the terminals of the concave resistors in Fig. 8-7*b* as shown in Fig. 8-7*d*.

By a dual argument, the convex DP plot in Fig. 8-7*e* can be realized by two convex resistors in series as shown in Fig. 8-7*f*. Again, notice that each time a new convex resistor with $I = i_3 > i_2$ is added in series, a new breakpoint is introduced and only the slope of the last segment of the original DP plot is altered. The complementary DP plot in Fig. 8-7*g* can be realized by the complementary network shown in Fig. 8-7*h*.

Consider next the DP plot shown in Fig. 8-7*i*. Notice that it is made up of a concave curve (segments 1, 2, 3) and a convex curve (segments 4, 5).[1] As usual, the first three segments can be realized by the two concave resistors shown in Fig. 8-7*j*. The remaining two segments are realized by the two convex resistors in the same figure. The parameters (G_j, E_j) of each concave resistor and (R_j, I_j) of each convex resistor are easily determined from the corresponding two adjacent segments. The basic procedure for realizing an arbitrary DP plot in the first quadrant should now be obvious. Starting from the leftmost segment, the DP plot is realized one segment at a time. Each new concave (convex) segment is realized by connecting a concave (convex) resistor in parallel (series) with the previously synthesized network. Observe that while the DP plot is being built up from left to right, the network is being built up from right to left. We shall refer to this procedure as the *segment-by-segment method.* It is important to observe that

[1] To be systematic, we shall always label the segments from the left to the right. Moreover, the voltage and current at the left breakpoint of segment j will be denoted by v_j and i_j. Similarly, the conductance and resistance of segment j will be denoted by g_j and r_j, respectively, where $g_j = 1/r_j$.



(b) $G_1 = g_1$ mhos, $E_1 = v_1$ volts
$G_2 = (g_2 - g_1)$ mhos, $E_2 = v_2$ volts

(d) $G_1 = g_1$ mhos, $E_1 = v_1$ volts
$G_2 = (g_2 - g_1)$ mhos, $E_2 = v_2$ volts

(f) $R_1 = r_1\,\Omega$, $I_1 = i_1$ amp
$R_2 = (r_2 - r_1)\,\Omega$, $I_2 = i_2$ amp

(h) $R_1 = r_1\,\Omega$, $I_1 = i_1$ amp
$R_2 = (r_2 - r_1)\,\Omega$, $I_2 = i_2$ amp

(j) $G_1 = g_2$ mhos, $E_1 = v_2$ volts; $G_2 = (g_3 - g_2)$ mhos, $E_2 = v_3$ volts
$R_3 = (r_4 - r_3)\,\Omega$, $I_3 = i_4$ amp; $R_4 = (r_5 - r_4)\,\Omega$, $I_4 = i_5$ amp

Fig. 8-7. Graphical illustration of how both concave and convex DP plots can be realized with the help of concave and convex resistors.

the key concept which makes this method so easy to apply is that each time a concave, or a convex, resistor is added the previously synthesized DP plot remains unchanged. Only the last segment is decomposed into two segments. Observe also that the segment-by-segment method requires as many concave and convex resistors as there are breakpoints.

So far, we have restricted our DP plot to the first quadrant (or the third quadrant).[1] Consider now an arbitrary monotonic DP plot. Our first step will be to apply the nonlinear Thévenin's theorem and obtain a new DP plot which passes through the origin. This is achieved simply by connecting a battery in series with the network (to be synthesized) having the new DP plot. Since the original DP plot is monotonic, it is clear that the new DP plot can lie only in the first and the third quadrants. Since the above preliminary step is rather trivial to do, there is no loss of generality in our restricting the following discussions only to passive monotonic DP plots. The segment-by-segment method for realizing this type of DP plot is best explained with the help of a numerical example.

EXAMPLE

Consider the passive monotonic DP plot shown in Fig. 8-8*a*. As usual, we label the segments from 1 to 8. The pertinent information describing each segment is tabulated in Fig. 8-8*b*. Again, the voltages e_j and i_j pertain to the left breakpoint of segment j. To synthesize this DP plot, the simplest procedure is to start with the segment through the origin, segment 5 in this case. The segments in the first quadrants are then synthesized one at a time. After this is done, we repeat the procedure to synthesize the segments in the third quadrant.[2]

Starting with segment 5, we obtain the resistor $R_5 = 2$ kΩ. Since segments 6 and 7 form parts of the concave curve, they can be realized by two concave resistors in parallel as shown in Fig. 8-8*c*. The parameters G_6 and E_6 are simply obtained from $G_6 = g_6 - g_5 = 1.5$ mmho and $E_6 = v_6 = 10$ volts. Similarly, $G_7 = g_7 - g_6 = \infty$ and $E_7 = v_7 = 12.5$ volts. To obtain segment 8, we connect a convex resistor with $R_8 = r_8 - r_7 = 4$ kΩ and $I_8 = i_8 = 15$ ma. The network to the right of terminals *a-b*, therefore, realizes the DP plot in the first quadrant.

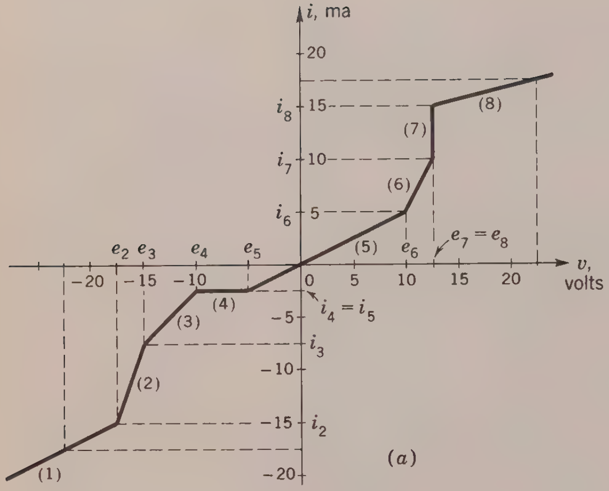
To synthesize the DP plot in the third quadrant, we first connect a convex resistor in series with the preceding network and obtain segment 4. The parameters of this resistor are given by $R_4 = r_4 - r_5 = \infty$ and $I_4 = -i_5 = 2.5$ ma.† The next two seg-

---

[1] Except portions of the DP plot which lie on the i axis or the v axis.

[2] Alternately, we can synthesize the DP plot in the third quadrant before going to the first quadrant. In fact, we may even jump back and forth, provided the order taken in each quadrant is consecutive.

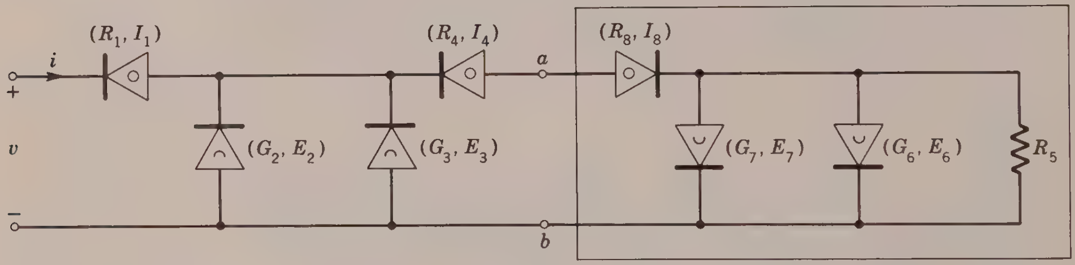
† The *negative* sign is needed to compute for the voltage and current parameters of all segments in the third quadrant because we are using the *complements* of the concave and the convex resistors, with the terminals transposed.



(a)

Segment number	g_j, mmhos	r_j, kΩ	e_j, volts	i_j, ma
(1)	0.5	2.0		
(2)	3.0	1/3	−17.5	−15
(3)	1.0	1.0	−15	−7.5
(4)	0	∞	−10	−2.5
(5)	0.5	2.0	−5	−2.5
(6)	2.0	0.5	10	5
(7)	∞	0	12.5	10
(8)	0.25	4.0	12.5	15

(b)



$R_5 = 2$ kΩ

(G_7, E_7): $G_7 = \infty$, $E_7 = 12.5$ volts

(R_4, I_4): $R_4 = \infty$, $I_4 = 2.5$ ma

(G_2, E_2): $G_2 = 2$ mmhos, $E_2 = 15$ volts

(G_6, E_6): $G_6 = 1.5$ mmhos, $E_6 = 10$ volts

(R_8, I_8): $R_8 = 4$ kΩ, $I_8 = 15$ ma

(G_3, E_3): $G_3 = 1$ mmhos, $E_3 = 10$ volts

(R_1, I_1): $R_1 = 5/3$ kΩ, $I_1 = 15$ ma

(c)

Fig. 8-8. A numerical example illustrating the general procedure for synthesizing any monotonic DP plot.

ments (3 and 2) are part of a complementary concave segment and hence can be realized by the two concave resistors shown in Fig. 8-8*c*. The parameters of these resistors are given, respectively, by $G_3 = g_3 - g_4 = 1.0$ mmho and $E_3 = -v_4 = 10$ volts, and by $G_2 = g_2 - g_3 = 2.0$ mmho and $E_2 = -v_3 = 15$ volts. Finally, to realize the last convex segment (segment 1), we connect a convex resistor with $R_1 = r_1 - r_2 = \frac{5}{3}$ kΩ and $I_1 = -i_2 = 15$ ma.

The above method is clearly applicable for realizing any passive monotonic DP plot. We shall now describe another method which is sometimes more convenient.

The decomposition method is based on the concept of splitting a DP plot into two simpler components. Each component DP plot can be further decomposed into even simpler components which are more amenable to synthesis. For example, if the DP plot contains a portion which is identical with the v-i curve of an existing device, such as a varistor, then it is more practical to decompose the DP plot into two component DP plots, one coinciding with the device's v-i curve.

Another technique for decomposition is motivated by the simplicity in synthesizing DP plots which lie only in the first quadrant. The technique consists of splitting the DP plot along the i axis or along the v axis. Accordingly, we have the *series-decomposition technique* and the *parallel-decomposition technique.*

To illustrate the series-decomposition technique, let us consider the same DP plot Γ in Fig. 8-8*a* and split it along the i axis, thereby obtaining two component plots Γ_1 and Γ_2. This first plot Γ_1 coincides with the DP plot Γ in the first quadrant, but with the negative i axis included as part of Γ_1, as shown by the v_a-i_a curve in Fig. 8-9*a*. The second plot Γ_2 is obtained by rotating that part of the DP plot Γ in the third quadrant by 180° with respect to the origin, and then adding the negative i axis as part of Γ_2, as shown by the v_b-i_b curve in Fig. 8-9*b*. Observe now that if we connect the networks N_a and N_b back to back in series as shown in Fig. 8-9*c*, the resulting DP plot will be exactly Γ. Since both Γ_1 and Γ_2 are now restricted to the first quadrants, they can be easily realized by the methods described earlier. Moreover, since only the first quadrant is of interest here, the junction diode normally present inside the concave and convex resistors can be eliminated. The portion of the v_a-i_a and v_b-i_b curves coinciding with the negative-current axis can be approximately realized by a single junction diode across the input terminals. For example, the two circuits shown in Fig. 8-9*d* and *e* are two simple realizations for Γ_1 and Γ_2, respectively.

The parallel-decomposition technique is the dual of the technique described above. This time, the DP plot is split along the voltage axis to obtain the v_a-i_a and v_b-i_b curves shown in Fig. 8-9*f* and *g*. Again, these DP plots are in the first quadrants and can be easily realized. The final DP plot is obtained by connecting the two networks N_a and N_b in the back-to-front parallel configuration shown in Fig. 8-9*h*.

One practical disadvantage of the segment-by-segment method is that the breakdown voltage E (current I) of each additional zener (constant-current) diode must be greater than that of the

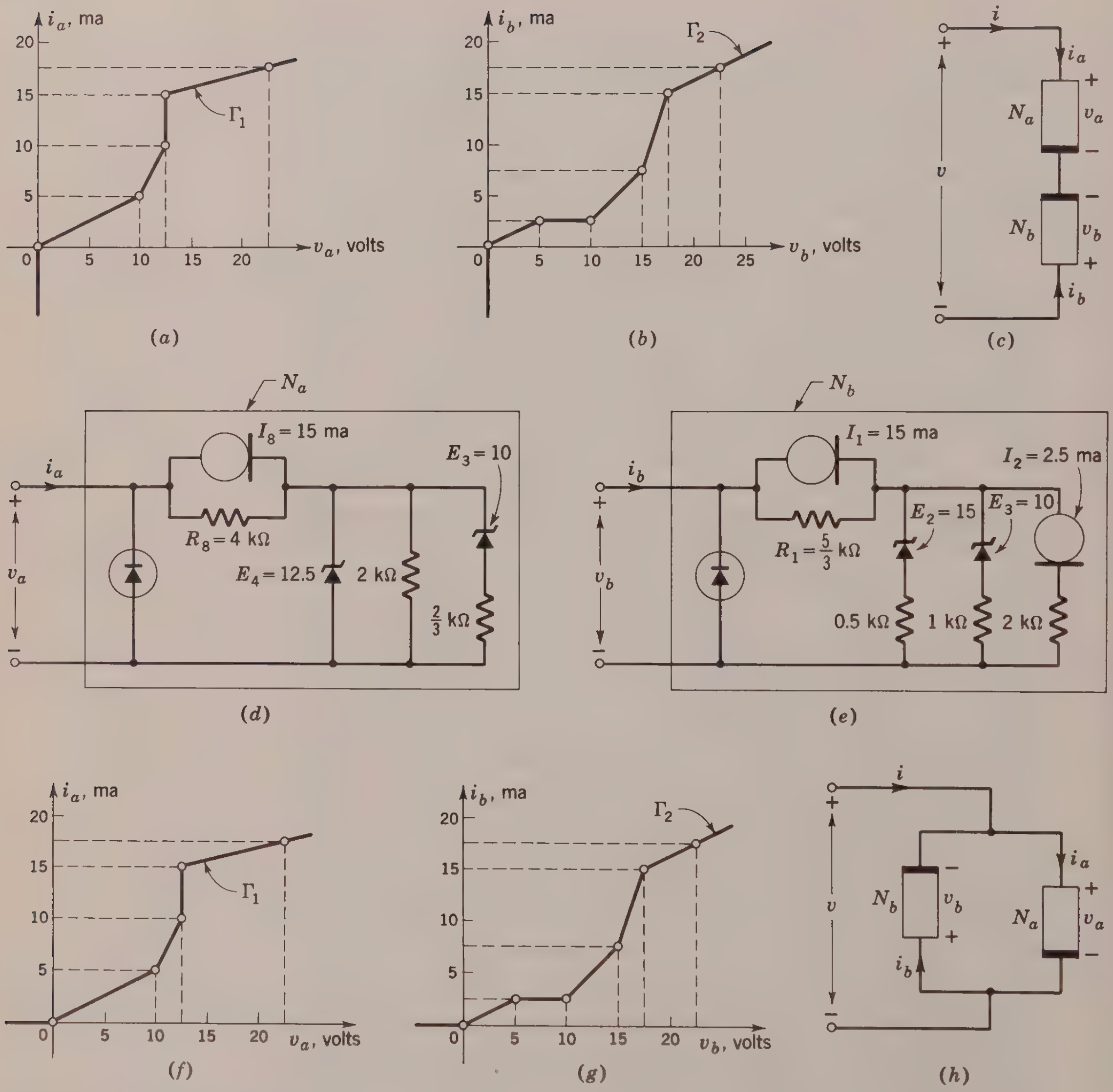
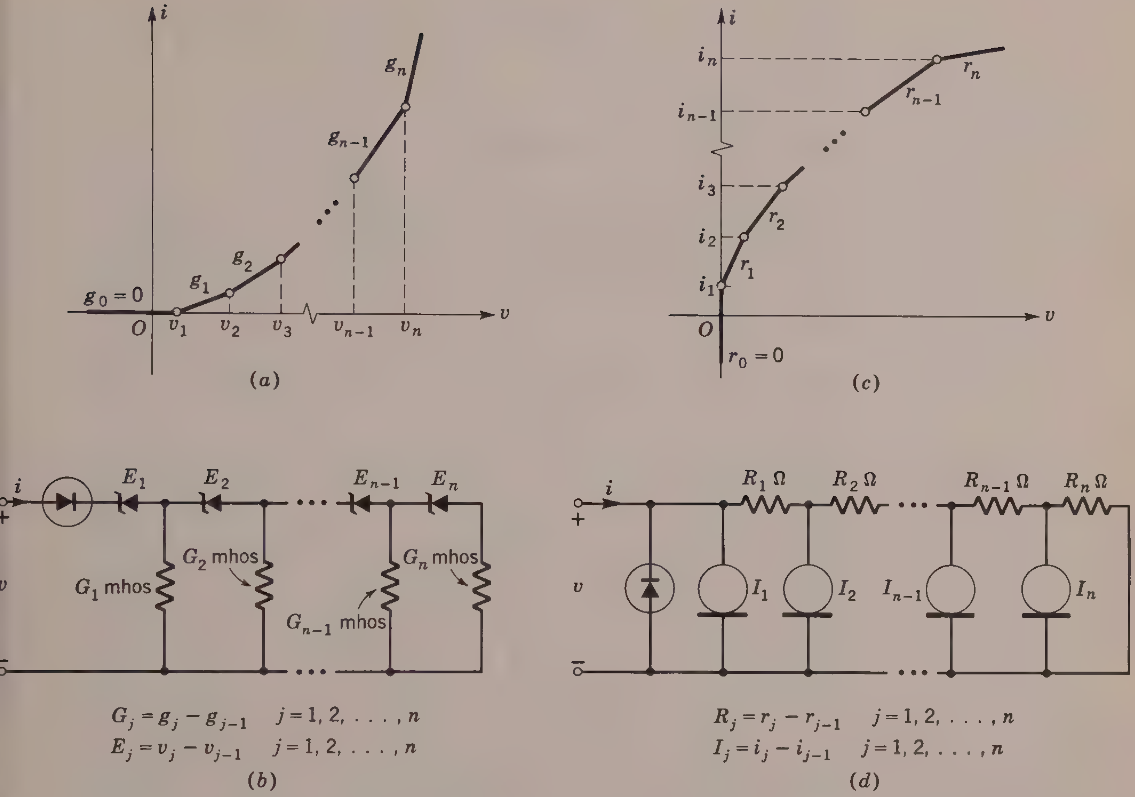


Fig. 8-9. Any passive monotonic DP plot can be realized by either the series-decomposition technique or the shunt-decomposition technique.

preceding diode. Since higher-voltage (current) zener (constant-current) diodes are more expensive,[1] we shall conclude this section by presenting two nonlinear ladder networks which do not have this problem. It can be easily verified that the concave DP plot shown in Fig. 8-10*a* can always be realized by the zener-diode ladder network shown in Fig. 8-10*b*. Notice that the breakdown voltage of each zener diode depends only on the voltage between

---

[1] This is because higher-voltage (current) zener (constant-current) diodes are usually realized in



$G_j = g_j - g_{j-1} \quad j = 1, 2, \ldots, n$

$E_j = v_j - v_{j-1} \quad j = 1, 2, \ldots, n$

$R_j = r_j - r_{j-1} \quad j = 1, 2, \ldots, n$

$I_j = i_j - i_{j-1} \quad j = 1, 2, \ldots, n$

Fig. 8-10. Two practical nonlinear ladder networks for realizing an arbitrary concave or convex DP plot in the first quadrant.

two adjacent breakpoints and not on the absolute voltage of each breakpoint as in the segment-by-segment method. By duality arguments, the convex v-i curve shown in Fig. 8-10*c* can be realized by the constant-current diode ladder network shown in Fig. 8-10*d*.[1]

Exercise 1: Explain why we choose the conductance (resistance) parameters g_j and G_j (r_j and R_j) when dealing with concave (convex) DP plots and resistors.

Exercise 2: Verify the following observations concerning the network shown in Fig. 8-8*c*. (*a*) The two-port subnetwork on each side of terminals a-b (excluding R_5) can be interchanged. (*b*) The junction diode associated with the concave and convex resistors to the right of terminals a-b can be replaced by a single diode inserted (in series) through terminal a.

Exercise 3: Show that the segments of the DP plot of Fig. 8-8*a* can be realized in the following orders: (*a*) segments 5, 4, 6, 3, 7, 2, 8, 1; (*b*) segments 5, 6, 7, 4, 3, 2, 8, 1; (*c*) segments 5, 6, 4, 3, 2, 7, 1, 8. Find a network realization for each case.

Exercise 4: Find a network realization for N_a and N_b of Fig. 8-9*h*.

---

practice by connecting several lower-voltage (current) zener (constant-current) diodes in series (parallel).

[1] The v-i curves of the zener diodes, constant-current diodes, and junction diodes in Fig. 8-10 are assumed to be as given in Table 8-2.

Exercise 5: Derive the equations specifying G_j, E_j for the zener-diode ladder network and R_j, I_j for the constant-current diode ladder network in Fig. 8-10.

8-3-2 SYNTHESIS OF NONMONOTONIC DP PLOTS

We have shown that any monotonic DP plot can be realized by using only passive components (top portion of Table 8-2) and at most one independent source such as a battery. To realize a nonmonotonic DP plot, however, active components are generally needed.

Consider first the current-controlled DP plot shown in Fig. 8-11*a*. Notice that if we connect a large-enough positive resistance (Fig. 8-11*b*) in series with a network N having this DP plot, we would obtain the monotonic DP plot shown in Fig. 8-11*c*. If a current-controlled DP plot contains *m negative-resistance segments,* and if the most negative segment[1] has a resistance $-R_{\max}$, then by choosing a positive resistance $R \geq R_{\max}$ in series with N, a monotonic DP can always be obtained. Suppose now we connect a positive resistance R and a negative resistance $-R$ in series with the network N as shown in Fig. 8-11*d*. Obviously, the DP plot of this network is identical with that of N (Fig. 8-11*a*). Notice, however, that the positive resistor R and the network N can be replaced by an equivalent network N' whose DP plot is monotonic. This DP plot can be realized by the techniques presented in the preceding section, and hence the DP plot shown in Fig. 8-11*a* can be realized by the network shown in Fig. 8-11*f*.

The dual technique can be applied to synthesize any voltage-controlled DP plot. This time, the two resistors must be connected in parallel, and the magnitude of the conductance must be chosen greater than the conductance of the most negative segment. Hence, the above discussions can be summarized by the following: *Any current-controlled or voltage-controlled DP plot can be realized by exactly one negative resistance and a network with a monotonic DP plot.*[2]

It must be emphasized, however, that in synthesizing nonmonotonic plots, unlike monotonic plots, which are always stable and easy to synthesize, the presence of the negative resistance may often lead to unwanted oscillations. The reason for the oscillation will be given in Chap. 14. Another problem associated with the above realization techniques is that the DP plot is extremely sensitive to changes in the value of the negative resistance. We shall now discuss another method which is usually more stable from a practical point of view. We shall present the method for a voltage-

---

[1] This corresponds to that negative-resistance segment which most closely approaches a horizontal line.

[2] Recall that a negative resistance can be realized in practice by connecting a positive resistance across one port of a negative-impedance converter (see Sec. 3-7).

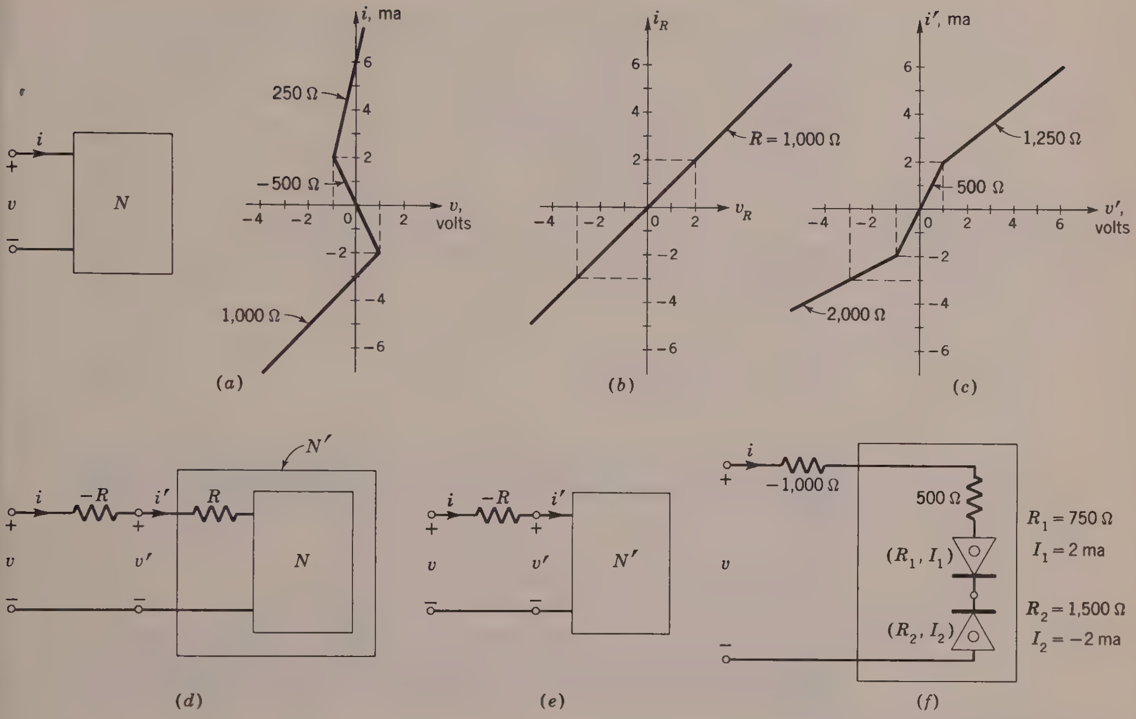


Fig. 8-11. An illustration showing that any current-controlled DP plot can be realized by connecting a negative resistor in series with a network characterized by a monotonic DP plot.

controlled DP plot, leaving the dual for the problems. The foundation of this method is based on the following two theorems.

SCALOR-ROTATOR REALIZATION THEOREM

Any passive voltage-controlled DP plot with finite slopes can be realized by a cascade connection of a scalor, a rotator, and a network with a monotonic DP plot.

In order to prove this theorem, let us represent the voltage-controlled DP plot by $i = f(v)$ and let its slope be given by $G(v) = df(v)/dv$. If a resistor with this DP plot is connected across port 2 of the rotator as in Fig. 3-27 in Chap. 3, we can derive the slope di_1/dv_1 of the new DP plot seen across port 1 of the rotator by first substituting $v_2 = v$ and $i_2 = -i = -f(v)$ into Eqs. (3-88) and (3-89);[1] thus

$$v_1 = (\cos\theta)v - (\sin\theta)f(v) \tag{8-23}$$

$$i_1 = (\sin\theta)v + (\cos\theta)f(v) \tag{8-24}$$

---

[1] For simplicity, we assume that the scale factor $R = 1$.

Differentiating these equations with respect to v, we obtain

$$\frac{dv_1}{dv} = \cos\theta - (\sin\theta)G(v) \tag{8-25}$$

$$\frac{di_1}{dv} = \sin\theta + (\cos\theta)G(v) \tag{8-26}$$

Dividing Eq. (8-26) by Eq. (8-25) and simplifying, we obtain the desired slope

$$\frac{di_1}{dv_1} = \frac{\tan\theta + G(v)}{1 - (\tan\theta)G(v)} \tag{8-27}$$

In order for the *rotated* v_1-i_1 curve to be monotonic, we require that $di_1/dv_1 \geq 0$. It is easy to see from Eq. (8-27) that this condition implies

$$-\tan\theta \leq G(v) \leq \cot\theta \tag{8-28}$$

If we let $G_{\min}$ and $G_{\max}$ be the minimum and maximum values of $G(v)$, respectively, then Eq. (8-28) implies that the range of the angle of rotation θ must be restricted by

$$\tan^{-1}(-G_{\min}) \leq \theta \leq \cot^{-1}(G_{\max}) \tag{8-29}$$

Depending on the given function $f(v)$, we may or may not be able to find a range of values for θ which satisfies Eq. (8-29). However, if we connect a current scalor with a scaling constant k_i in cascade with the rotator, then the slope of the original v-i curve should be divided by k_i. In this case, Eq. (8-29) becomes

$$\tan^{-1}\frac{-G_{\min}}{k_i} \leq \theta \leq \cot^{-1}\frac{G_{\max}}{k_i} \tag{8-30}$$

It is now easy to see that if the value of k_i is chosen such that

$$k_i > \sqrt{|G_{\min}G_{\max}|} \tag{8-31}$$

then since the slope of the DP plot is finite by hypothesis, the quantity under the radical sign is a finite number; therefore, $k_i \neq \infty$, and there is always a range of values for θ which satisfies Eq. (8-30).

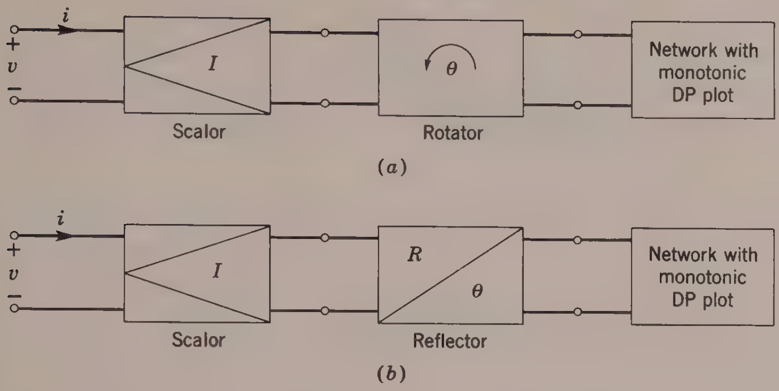

i
+
v
−
I
Scalor
θ
Rotator
Network with monotonic DP plot
(a)
i
+
v
−
I
Scalor
R
θ
Reflector
Network with monotonic DP plot
(b)

Fig. 8-12. Two basic circuits utilizing a scalor, a rotator, and a reflector for realizing arbitrary voltage-controlled DP plots.

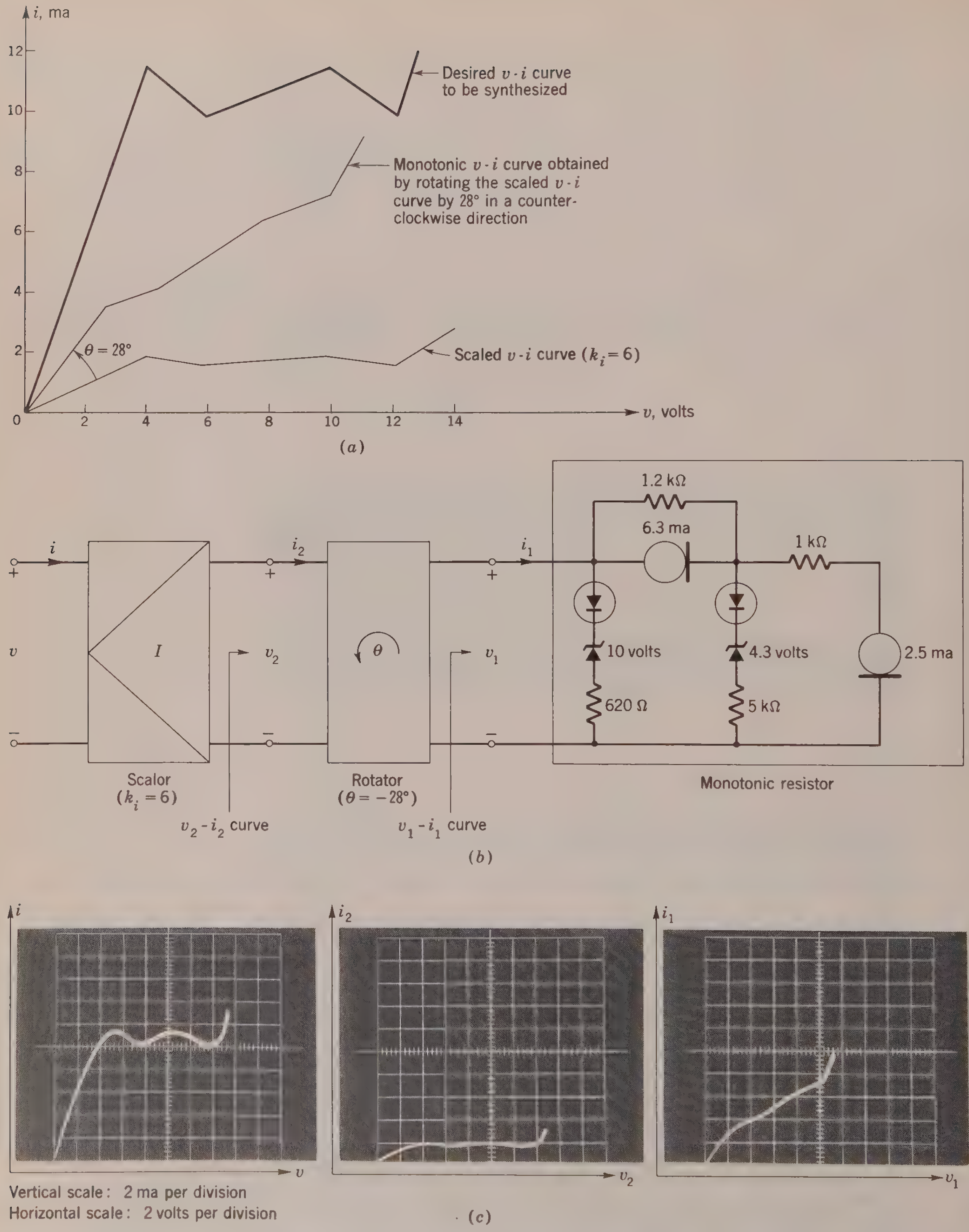
What we have shown, therefore, is that any voltage-controlled v-i curve can be transformed into a monotonic v_1-i_1 curve by simply choosing the value k_i of the current scalor and the angle of rotation θ of the rotator in accordance with Eqs. (8-30) and (8-31), respectively. If we now reverse the steps in the above proof, we shall arrive at the procedure for synthesizing a voltage-controlled DP plot by the scalor-rotator network shown in Fig. 8-12*a*. This proves the theorem. Before we consider a numerical example, we shall state a somewhat analogous result using a reflector instead of a rotator.

SCALOR-REFLECTOR REALIZATION THEOREM

Any passive voltage-controlled DP plot with finite slopes can be realized by a cascade connection of a scalor, a reflector, and a network with a monotonic DP plot.

The proof of this theorem follows arguments similar to those given earlier and will, therefore, be left as an exercise. The circuit configuration for this case is shown in Fig. 8-12*b*.

As a numerical example, consider the voltage-controlled v-i curve shown in Fig. 8-13*a*. It is easy to see that the value $k_i = 6$ satisfies Eq. (8-31), and with this value the angle $\theta = 28°$ is chosen [arbitrarily, so long as it satisfies Eq. (8-30)]. Hence, the current (ordinate) of the v-i curve in Fig. 8-13*a* is divided by the number $k_i = 6$ as shown by the lower dotted curve. If we rotate this curve by 28° counterclockwise, we obtain, indeed, the *monotonic* DP plot shown in Fig. 8-13*a*. This DP plot is then synthesized by the segment-by-segment method. The resulting circuit is shown connected to the scalor-rotator combination as in Fig. 8-13*b*. In order



Vertical scale: 2 ma per division
Horizontal scale: 2 volts per division

Fig. 8-13. A numerical example illustrating the synthesis of a voltage-controlled v-i curve by the scalor-rotator combination.

to compare the way the actual curve differs from the prescribed curve, the network in Fig. 8-13*b* was built, and the v_1-i_1, v_2-i_2, and v-i curves are measured as shown in Fig. 8-13*c*.[1] These scope tracings show a remarkable agreement with the curves in Fig. 8-13*a*.

Exercise 1: Show that any voltage-controlled DP plot can be realized by a negative resistor and a network with a monotonic DP plot.

Exercise 2: Derive Eqs. (8-29), (8-30), and (8-31).

Exercise 3: Prove the scalor-reflector realization theorem and verify that the constant k_i of the scalor and the angle of reflection θ are also given by Eqs. (8-31) and (8-30), respectively.

Exercise 4: State and prove the dual theorems for realizing a current-controlled DP plot by means of scalors, rotators, and reflectors.

8-4 SYNTHESIS OF TC PLOTS

We have already seen a number of practical applications (for example, Sec. 6-6) which require the synthesis of a network with a prescribed TC plot. We shall see more applications in the next two chapters. Since there are four different types of TC plots, we shall describe at least one method for each.

Synthesis of i_o-vs.-v_i TC plot This type of TC plot can be realized by the simple series circuit shown in Fig. 8-14*a*. The DP plot of the nonlinear resistor R is identical with the prescribed TC plot but with the i_o axis replaced by the i axis. This follows from the observation that $v = v_i$ and $i = i_o$. The problem of realizing a prescribed i_o-vs.-v_i TC plot reduces, therefore, to the synthesis of one DP plot, a task we already know how to do.

Synthesis of v_o-vs.-i_i TC plot This type of TC plot is the dual of that presented above. Hence, the dual circuit shown in Fig. 8-14*b* constitutes a simple realization. The DP plot of R in this case is identical with the prescribed TC plot but with the v_o axis replaced by the v axis.

Synthesis of v_o-vs.-v_i TC plot This type of TC plot can be realized by the circuit shown in Fig. 8-14*c*. The DP plot of the nonlinear resistor is identical with the prescribed TC plot but with the v_o axis replaced by the i axis. Observe that this circuit is obtained by adding a current-controlled voltage source to the circuit of Fig. 8-14*a*.

---

[1] The "origin" in these tracings is located at the bottom left corner.

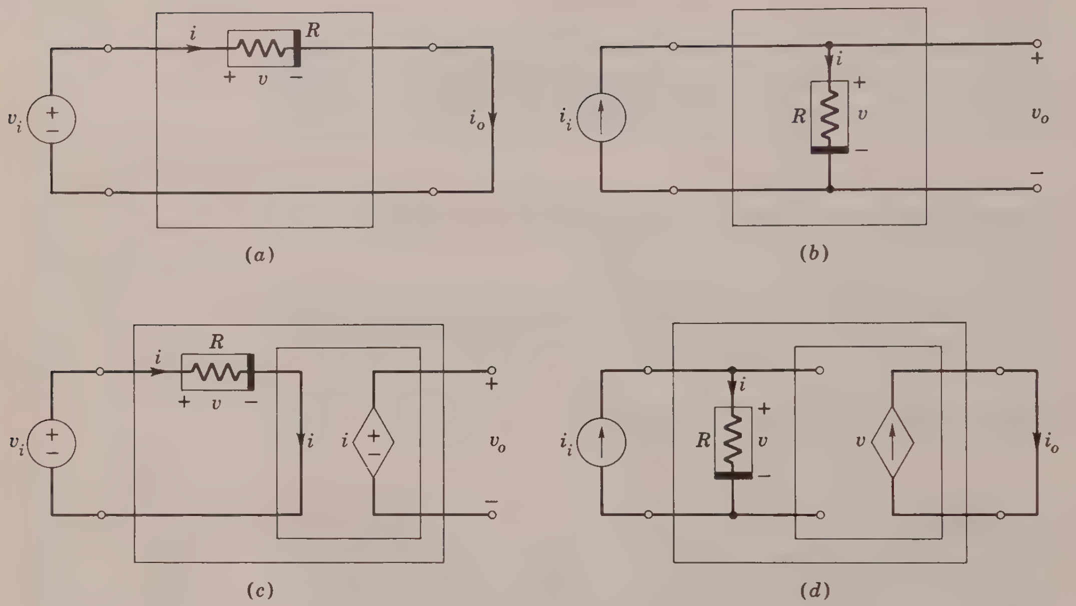


Fig. 8-14. Basic circuit realizations for four different types of TC plots.

Synthesis of i_o-vs.-i_i TC plot This type of TC plot is the dual of that presented above and can, therefore, be realized by the dual circuit shown in Fig. 8-14*d*. This time, the DP plot of the nonlinear resistor R is identical with the prescribed TC plot but with the i_o axis replaced by the v axis.

While it is true that the last two realization techniques are completely general, they are not necessarily the most practical because of the need for the controlled sources.[1] On many occasions it may be desirable to choose a more practical circuit. We shall now describe two methods for achieving this.

8-4-1 TC PLOT SYNTHESIS BY A NONLINEAR VOLTAGE- OR CURRENT-DIVIDER NETWORK

We shall now show that any v_o-vs.-v_i TC plot can be realized by a nonlinear voltage-divider network and any i_o-vs.-i_i TC plot can be realized by a nonlinear current-divider network. Consider first the nonlinear voltage-divider network shown in Fig. 8-15*a*. The resistor R_1 is linear, and its resistance may be arbitrarily chosen.[2] Our problem is to find the appropriate DP plot for the nonlinear resistor R_2 in order to realize the prescribed TC plot. This can be easily found by applying the template method discussed earlier in Sec. 6-5-1. As will be obvious soon, it is more convenient to redraw

[1] Controlled sources are presently expensive to build and are more sensitive to parameter variations. These disadvantages, however, are gradually disappearing because of the advances in integrated-circuit technology.

[2] One may take advantage of this arbitrariness to select R_1 so that it includes the internal resistance of the "physical" voltage source v_i.

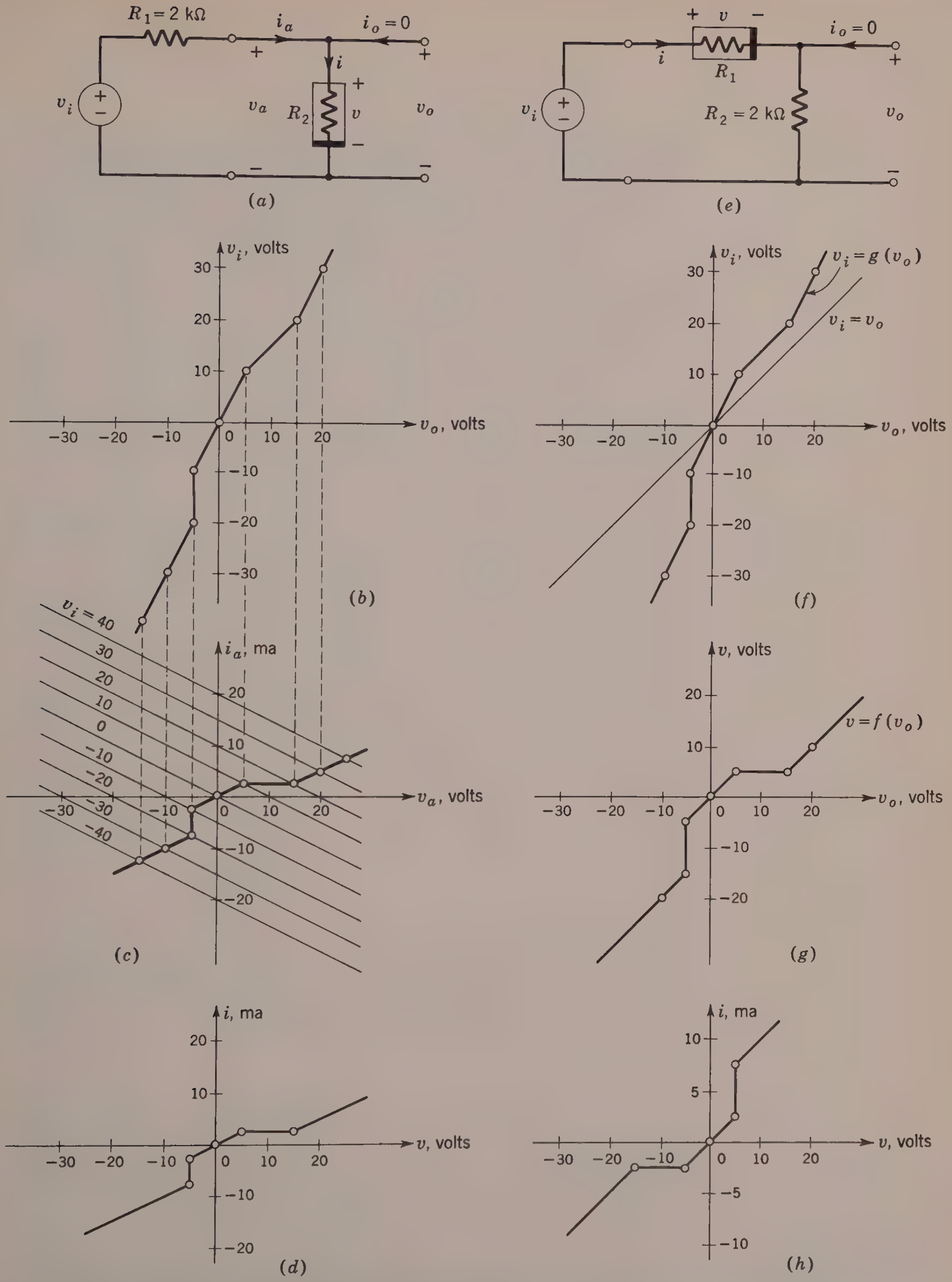


Fig. 8-15. Examples showing the graphical procedure for realizing a TC plot by two basic nonlinear voltage-divider networks.

the prescribed TC plot with v_i as the vertical axis and v_o as the horizontal axis. To illustrate the procedure, suppose the TC plot shown in Fig. 8-15*b* is to be synthesized. Suppose further that $R_1 = 2\ \text{k}\Omega$. Our first step is to draw the family of straight lines representing the v_a-i_a curve with v_i as the parameter as shown in Fig. 8-15*c*. The slope of these lines is negative because we have defined the reference direction for i_a in such a way that $i_a = i$. Notice that the i_a axis is aligned with the v_i axis of Fig. 8-15*b* so that the abscissa of each point on the TC plot is identical with the corresponding point in Fig. 8-15*c*. Moreover, since the given TC plot is piecewise-linear, we need only to drop a vertical guideline from each breakpoint until it intersects the line corresponding to the same value of v_i. Hence, the TC plot in Fig. 8-15*b* is easily "transferred" into the v_a-i_a plane as shown in Fig. 8-15*c*. Since $i_a = i$ and $v_a = v$, this *transferred* TC plot can be redrawn in the v-i plane as shown in Fig. 8-15*d*. This is the desired DP plot for R_2 and can be realized by the methods discussed earlier.

An alternate method for realizing an arbitrary TC plot is the voltage-divider network shown in Fig. 8-15*e*. This time the resistor R_2 is linear and may assume arbitrary values. Our problem is to find the appropriate DP plot for R_1. As usual, it is more convenient to plot the TC plot with v_i as the vertical axis as in Fig. 8-15*f*. If we denote this TC plot by $v_i = g(v_o)$, then the voltage v across R_1 is given by $v = v_i - v_o = g(v_o) - v_o$, or

$$v = f(v_o) \tag{8-32}$$

where

$$f(v_o) \equiv g(v_o) - v_o \tag{8-33}$$

This function represents a curve in the v-vs.-v_o plane obtained by subtracting the TC plot $v_i = g(v_o)$ from the line $v_i = v_o$ as shown in Fig. 8-15*g*.[1] The next step is to observe that

$$v_o = R_2 i \tag{8-34}$$

Substituting Eq. (8-34) for v_o in Eq. (8-32), we obtain the desired DP plot

$$v = f(R_2 i) \tag{8-35}$$

Notice that Eq. (8-35) is obtained simply by dividing the scale of the v_o axis in Fig. 8-15*g* by R_2, and then relabeling this axis by i.

---

[1] Although we denote Eq. (8-33) by a functional notation, the method is applicable even if $f(v_o)$ represents a multivalued curve.

The last step is to redraw this v-i curve with i as the vertical axis in order to conform with our usual practice. Hence, to obtain the DP plot of R_1, we first divide the scale of the v_o axis in Fig. 8-15*g* by $R = 2K$ and relabel this axis by i in milliamperes. The resulting DP plot for R_1 is shown in Fig. 8-15*h*.

It can be easily shown that if the slope of the TC plot $dv_o/dv_i \leq 1$ or if $dv_i/dv_o \geq 1$, then the DP plot for the nonlinear resistor R_2 in Fig. 8-15*a*, or the nonlinear resistor R_1, in Fig. 8-15*e*, will always be monotonic. For this class of TC plot, therefore, the above voltage-divider method is much more practical.

An important point that we have tacitly assumed so far is that there are no *loading effects;* that is, $i_o = 0$. In practice, the output terminals are invariably connected across an external load which may draw some current. In most cases, this external load can be represented by an equivalent load resistor R_L. If R_L is nonlinear, the circuit in Fig. 8-15*a* is more appropriate because the resistor R_L can be *included* as part of R_2 before the network is synthesized. For example, if the DP plot of R_2 (obtained without loading effect) is given by $i = g(v)$, and if the load resistor has a v_L-i_L curve given by $i_L = h(v_L)$, then since R and R_L are in parallel, we have $v = v_L$. Hence, we may "extract" this v_L-i_L curve from the v-i curve and obtain a new v'-i' curve given by $i' = g(v') - h(v') \equiv f(v')$. In other words, the DP plot of R_2 that must be synthesized is given by $i' = f(v')$ and not $i = g(v)$. On the other hand, if R_L is linear, then the circuit in Fig. 8-15*b* may be more convenient to use because in this case R_L can simply be chosen equal to R_2. In both cases the loading effect can be easily taken care of, provided the load remains unchanged.

Although our discussions so far have been restricted to the synthesis of v_o-vs.-v_i TC plots by voltage dividers, it is clear that the dual procedures can be used to synthesize any i_o-vs.-i_i TC plots by current dividers. We shall leave the details for doing this as an exercise.

8-4-2 TC PLOT SYNTHESIS BY OPERATIONAL AMPLIFIER CIRCUITS

The disadvantage of the voltage- and current-divider methods is that whenever the slope of the TC plot $dv_o/dv_i > 1$ or $di_o/di_i > 1$, the DP plot of the nonlinear resistor becomes nonmonotonic. The simplicity of these methods is thereby overshadowed by the need to synthesize nonmonotonic DP plots. We shall now present another method which does not have this disadvantage.

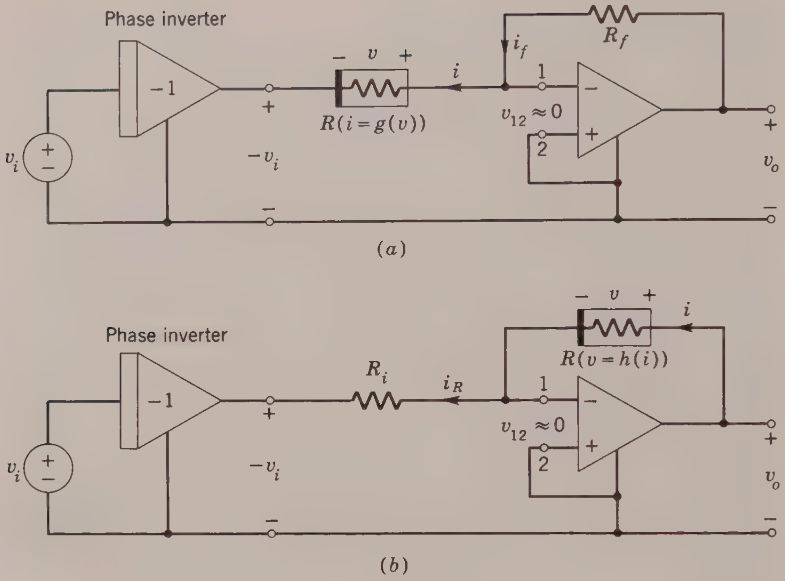


Fig. 8-16. Any v_o-vs.-v_i TC plot can be realized by an operational amplifier circuit using one nonlinear resistor.

Consider first the operational amplifier circuit shown in Fig. 8-16*a*. In view of the principle of virtual short circuit (Sec. 3-6-2), we obtain $i_f = i$ and

$$v_o = R_f i_f = R_f i = R_f g(v) \tag{8-36}$$

where $g(v)$ is the v-i curve of the nonlinear resistor R. But $v = -(-v_i) = v_i$, and hence Eq. (8-36) becomes

$$v_o = R_f g(v_i) \tag{8-37}$$

This equation shows that any TC plot $v_o = f(v_i)$ can be realized by the circuit in Fig. 8-16*a*, provided that the DP plot of the nonlinear resistor R is chosen to be

$$i = \frac{f(v)}{R_f} \tag{8-38}$$

Observe that if we replace R_f in Eq. (8-38) by a constant K_0, then this constant can be used to change the scale of the output voltage v_o of the TC plot. Observe also that Eq. (8-38) implies that the DP plot of the nonlinear resistor will be monotonic as long as the TC plot is monotonic.

An alternate realization using an operational amplifier is shown in Fig. 8-16*b*. If we denote the *v-i* curve of the nonlinear resistor R by $v = h(i)$, then the relation between v_o and v_i is easily found to be given by

$$v_o = h\left(\frac{v_i}{R_i}\right) \tag{8-39}$$

This equation implies that any TC plot $v_o = f(v_i)$ can be realized by the circuit shown in Fig. 8-16*b*, provided the DP plot of the nonlinear resistor R is chosen to be

$$v = f(R_i i) \tag{8-40}$$

Observe that if we replace R_i in Eq. (8-40) by a constant K_i, then this constant can be used to change the scale of the input voltage v_i of the TC plot. Again, the DP plot of the nonlinear resistor will be monotonic as long as the TC plot is monotonic.

If the TC plot is nonmonotonic, then all methods presented so far will require the synthesis of a nonmonotonic DP plot. Under appropriate conditions, however, it is possible to trade one nonmonotonic resistor with four monotonic ones, as will be demonstrated in the following section.

Exercise 1: Verify the realizations shown in Fig. 8-15*d* and *h* by actually constructing the TC plots.

Exercise 2: Show that the DP plot for the nonlinear resistor in each of the voltage-divider networks is monotonic if the slope of the TC plot $dv_o/dv_i \leq 1$ or if the reciprocal slope $dv_i/dv_o \geq 1$.

Exercise 3: There are applications wherein not only the TC plot and the load resistor are specified, but also the DP plot across the voltage-source terminals. This can be achieved by making both resistors in the voltage-divider network nonlinear. (*a*) Describe a general synthesis procedure for achieving this set of specifications. (*b*) Specialize the procedure to the case when the input DP plot is a straight line. Such a network is called a constant-resistance network.

Exercise 4: Describe the general procedure for realizing any i_o-vs.-i_i TC plot by means of a nonlinear current-divider network. Consider two cases, first with R_1 assumed to be linear, and then with R_2 assumed to be linear.

Exercise 5: Compare the advantages and disadvantages of the two circuits shown in Fig. 8-16.

8-5 SYNTHESIS OF JOINTLY PRESCRIBED DP PLOT, TC PLOT, AND LOAD v-i CURVE

The methods presented so far require the synthesis of a nonmonotonic DP plot whenever the TC plot contains one or more segments with negative slopes.[1] The fact that the TC plot of the trivial network consisting of a pair of crossed wires is given by $v_o = -v_i$ suggests that this requirement is not necessary in general. In fact, we shall now derive a method based on the lattice network shown in Fig. 8-17*a*. The foundation of this method is given by the following theorem.

LATTICE NETWORK SYNTHESIS THEOREM

A symmetric lattice network containing only monotonic resistors can always be synthesized to have a prescribed TC plot $v_o = T(v)$, a prescribed load resistor v_L-i_L curve $i_L = g_L(v_L)$, and a prescribed DP plot $i = g(v)$ across the input terminals provided the following three conditions are satisfied:

$$\text{Condition 1:} \qquad \frac{dg(v)}{dv} > 0 \qquad \text{and} \qquad \frac{dg_L(v_L)}{dv_L} > 0$$

$$\text{Condition 2:} \qquad -1 < \frac{dT(v)}{dv} < 1 \tag{8-41}$$

$$\text{Condition 3:} \qquad \left|\frac{dT(v)}{dv}\right| < \frac{dg(v)/dv}{dg_L(v_L)/dv_L}$$

Observe that condition 1 implies that both the DP plot and the load resistor v_L-i_L curve are strictly monotonically increasing functions. Condition 2 implies that the magnitude of the slope of the TC plot must be less than unity. The last condition imposes a constraint between the slope of the TC plot and the slopes of the DP plot and the v_L-i_L curve at corresponding points. If the curves are piecewise-linear, which we shall assume in the following discussion for simplicity, the slope at each breakpoint can be chosen to correspond to either one of the two segments.

The proof of this theorem will be given in Chap. 12. Meanwhile, let us assume that it is true and proceed to derive a synthesis procedure. Since the above theorem guarantees that all resistors of the lattice network will come out monotonic, it follows from the uniqueness theorem (Sec. 5-6) that the network has a unique solution. Hence, we may apply the symmetric lattice network theorem

---

[1] The nonmonotonic DP plot is necessary in the method presented earlier because these networks have common input and output terminals. In order to obtain a TC plot with negative-slope segments and using only monotonic resistors, the network topology must have more than one path for currents to flow to the output terminals.

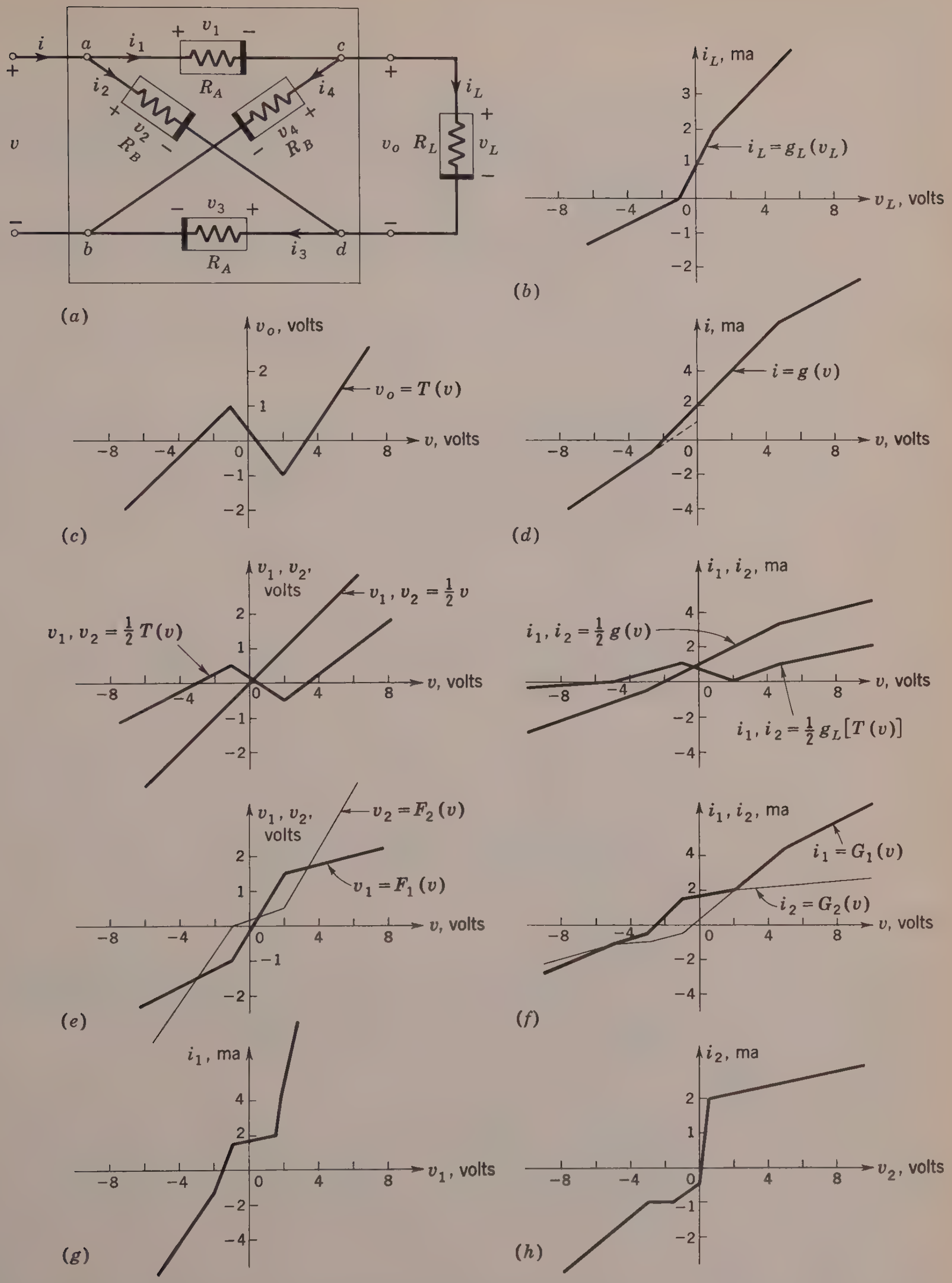


Fig. 8-17. Graphical procedures for synthesizing a nonlinear lattice network with a prescribed DP plot, TC plot, and load *v-i* curve.

(Sec. 7-7-1) to the network in Fig. 8-17*a* and conclude that $v_4 = v_2$ and $i_4 = i_2$. Suppose now we write the KVL equations around the loops formed by nodes *a-c-b* and *c-a-d* and use the relation $v_4 = v_2$ to obtain

$$v = v_1 + v_4 = v_1 + v_2$$

$$v_L = -v_1 + v_2$$

Solving for v_1 and v_2 from these equations, we obtain

$$v_1 = \tfrac{1}{2}(v - v_L) \tag{8-42}$$

$$v_2 = \tfrac{1}{2}(v + v_L) \tag{8-43}$$

Next, let us write the KCL equations at nodes *a* and *c* and use the relation $i_4 = i_2$ to obtain

$$i = i_1 + i_2$$

$$i_L = i_1 - i_4 = i_1 - i_2$$

Solving for i_1 and i_2 from these equations, we obtain

$$i_1 = \tfrac{1}{2}(i + i_L) \tag{8-44}$$

$$i_2 = \tfrac{1}{2}(i - i_L) \tag{8-45}$$

Let us now observe from Fig. 8-17*a* that since $v_L = v_o$, the TC plot $v_o = T(v)$ can also be written as

$$v_L = T(v) \tag{8-46}$$

Hence, the load resistor v_L-i_L curve $i_L = g_L(v_L)$ can be expressed in terms of the driving-point variable v by a *graphical composition* between this curve and the TC plot, namely,

$$i_L = g_L(T(v)) \tag{8-47}$$

Finally, let us substitute Eqs. (8-46) and (8-47) into Eqs. (8-42) to (8-45) to obtain

$$v_1 = \tfrac{1}{2}v - \tfrac{1}{2}T(v) \equiv F_1(v) \tag{8-48}$$

$$v_2 = \tfrac{1}{2}v + \tfrac{1}{2}T(v) \equiv F_2(v) \tag{8-49}$$

$$i_1 = \tfrac{1}{2}g(v) + \tfrac{1}{2}g_L(T(v)) \equiv G_1(v) \tag{8-50}$$

$$i_2 = \tfrac{1}{2}g(v) - \tfrac{1}{2}g_L(T(v)) \equiv G_2(v) \tag{8-51}$$

Observe that these equations are now functions of the same variable v and can be readily obtained graphically.

It remains for us to eliminate the variable v from the two curves $F_1(v)$ and $G_1(v)$ to obtain the v_1-i_1 curve of resistor R_A. This again can be readily done graphically. Similarly, we can graphically eliminate the variable v from the two curves $F_2(v)$ and $G_2(v)$ to obtain the v_2-i_2 curve of resistor R_B. Let us illustrate this procedure by an example.

EXAMPLE

Suppose the prescribed load resistor curve $i_L = g_L(v_L)$, the prescribed TC plot $v_o = T(v)$, and the prescribed DP plot $i = g(v)$ are given, respectively, in Fig. 8-17*b* to *d*. An inspection of these curves shows that conditions 1, 2, and 3 are satisfied, and hence, we are guaranteed that the method is valid.[1]

The first step in the synthesis procedure is to obtain the curves $v_1 = F_1(v)$ and $v_2 = F_2(v)$ as defined by Eqs. (8-48) and (8-49). This is easily done graphically as shown in Fig. 8-17*e*. Notice that the ordinates of the TC plot must be divided by 2 before carrying out the subtraction and addition operations graphically.

The next step is to obtain the curves $i_1 = G_1(v)$ and $i_2 = G_2(v)$ as defined by Eqs. (8-50) and (8-51). Before we do this, however, we must first obtain the curve $i_L = g_L(T(v))$ by a graphical composition between the load resistor curve $i_L = g_L(v_L)$ and the TC plot $v_L = T(v)$. This curve is shown in the upper portion of Fig. 8-17*f* together with the curve $i_1 = i_2 = \frac{1}{2}g(v)$ (obtained by dividing the ordinates of the DP plot by 2). The results of adding and subtracting the ordinates of these two curves are shown by the curves $i_1 = G_1(v)$ and $i_2 = G_2(v)$ in Fig. 8-17*f*.

The last step in the synthesis procedure is to obtain the v_1-i_1 curve of resistor R_A and the v_2-i_2 curve of resistor R_B.† The v_1-i_1 curve is easily obtained by a graphical elimination procedure between the curves $v_1 = F_1(v)$ of Fig. 8-17*e* and $i_1 = G_1(v)$ of Fig. 8-17*f*. The result is shown in Fig. 8-17*g*. Similarly, the v_2-i_2 curve is obtained by a graphical elimination procedure between the curves $v_2 = F_2(v)$ of Fig. 8-17*e* and $i_2 = G_2(v)$ of Fig. 8-17*f*. The result is shown in Fig. 8-17*h*. As expected, both resistors R_A and R_B are monotonic. They can, therefore, be easily realized by the methods described earlier.

The above graphical procedures are relatively simple. However, if a numerical error is committed, it is sometimes difficult to detect. A computer program can be easily written to carry out the graphical procedures above. Among other things, the computer can be instructed to check the three conditions given in Eq. (8-41) and to print out the complete solution if these conditions are satisfied. If one or more conditions are not satisfied, the computer can still be instructed to carry out the procedures, only this time we expect the resistors R_A and R_B to be nonmonotonic and hence the solutions may or may not be correct. In this case, the computer must be instructed to find the DP and TC plots of the synthesized network and to check them against the specifications.

---

[1] Unlike the other methods which are applicable to all cases, although we may end with nonmonotonic resistors, the present method is not valid in general if the three conditions are not satisfied.

† Since the lattice network is symmetrical, the v_1-i_1 curve of the upper resistor R_A is identical with the v_3-i_3 curve of the lower resistor. Similarly, the v_2-i_2 curve of resistor R_B across terminals *a-d* is identical with the v_4-i_4 curve of resistor R_B across terminals *c-b*.

There is nothing we can do if the conditions are not satisfied. In fact, as we shall show in Chap. 12, if the magnitude of the slope of the TC plot is greater than 1 at some points on the plot, then the resistors will be nonmonotonic. Physically, this agrees with our intuition that it is not possible for a network containing only monotonic resistors to produce a larger output voltage than the applied input voltage. This is because some of the input power will be dissipated as heat.

If we are interested only in the TC plot, however, we can divide its ordinate by an appropriate constant so that the three conditions are always satisfied. In this case, we might as well assume that the load resistor v_L-i_L curve and the DP plot are both linear. This TC plot can then be synthesized by the above lattice network. The final TC plot is obtained by connecting a voltage-controlled voltage source in cascade with the resulting lattice network. Hence, we have shown that any $v_o = T(v)$ TC plot can be realized by a symmetric lattice network (containing only monotonic resistors) and possibly a voltage-controlled voltage source.

Exercise 1: Verify that the network synthesized in Fig. 8-17 indeed satisfies the specifications. HINT: Apply the method used in Example 2 of Sec. 7-6.

Exercise 2: In many practical applications, the load resistor v_L-i_L curve and the DP plot are both linear. Show how the graphical procedures can be simplified in this case. Apply this procedure to realize the TC plot shown in Fig. 8-17*c* but with the load resistor and DP plot both represented by a 1-kΩ linear resistance.

Exercise 3: Find the procedures for synthesizing a prescribed TC plot $i_o = T(i)$, a prescribed load resistor curve $i_L = g_L(v_L)$, and a prescribed DP plot $i = g(v)$ by the symmetric lattice network shown in Fig. 8-17*a*. What are the conditions that must be satisfied in this case?

Exercise 4: Repeat Exercise 3 but with the TC plot replaced by $v_o = T(i)$ and then $i_o = T(v)$.

8-6 SUMMARY

Synthesis of operating points (biasing)

1. Any two-terminal resistor can be biased to operate at a prescribed point on its v-i curve by a resistor-battery *biasing network*. To satisfy other design criteria, a nonlinear biasing network may be desirable.
2. Any three-terminal resistor can be biased to operate at a prescribed point on its characteristic curves by one of the three biasing networks tabulated in Table 8-1.
3. The thermal-runaway phenomenon may occur in both two-terminal and three-terminal resistors. This phenomenon can be avoided by a properly designed biasing network.

Synthesis of DP plots

1. Any monotonic DP plot can be realized by using only linear resistors, concave resistors, convex resistors, and, at most, one battery. The concave and convex resistors can, in turn, be realized by junction diodes, zener diodes, constant-current diodes, and linear resistors. The synthesized network is always stable. The two methods are the following:
   a. Segment-by-segment method.
   b. Series- and parallel-decomposition methods. In these methods, the concave portion of the DP plot can be easily realized by a nonlinear ladder network containing only zener diodes, linear resistors, and one junction diode. The convex portion can be realized by a nonlinear ladder network containing only constant-current diodes, linear resistors, and one junction diode.
2. Any current- or voltage-controlled DP plot can be realized by three methods:
   a. A negative resistor in series (parallel) with a monotonic resistor
   b. A scalor-rotator combination and a monotonic resistor
   c. A scalor-reflector combination and a monotonic resistor

Synthesis of TC plots

1. Any one of the four types of TC plots can be realized by a nonlinear resistor and a controlled source.
2. Any monotonic TC plot is best realized by the following more practical methods:

a. If the slope of each point on the v_o-vs.-v_i (i_o-vs.-i_i) TC plot is positive and less than unity, use the nonlinear voltage-divider (current-divider) method.
b. If the slope of each point on the v_o-vs.-v_i TC plot is greater than unity, use the operational amplifier method.
c. If the magnitude of the slope of each point on the v_o-vs.-v_i (i_o-vs.-i_i) TC plot is less than unity (but may be negative), use the symmetric lattice network.
d. If the magnitude of the slope of the v_o-vs.-v_i (i_o-vs.-i_i) TC plot is greater than unity (but may be negative), a voltage-controlled voltage source (current-controlled current source) may be used in cascade with the symmetric lattice network.

Synthesis of jointly prescribed DP plot, TC plot, and load v-i curve This set of simultaneous specifications can be realized by a symmetric lattice network under appropriate conditions. (For the v_o-vs.-v_i TC plots, the conditions are given by Eq. 8-41.)

PROBLEMS

8-1 The biasing-circuit configuration 3 in Table 8-1 is to be used to bias a *p-n-p* transistor type 2NXX18B. The desired operating point is given by $I_B = -0.2$ ma and $V_{CB} = -5$ volts. The following biasing-circuit elements are chosen: $E = -10$ volts, $R_4 = 50\ \text{k}\Omega$, $R_2 = 1.5\ \text{k}\Omega$. Find the other element values of the biasing circuit.

8-2 The biasing-circuit configuration 2 in Table 8-1 is to be used to bias a p-channel FET type 2NXX20. The operating point is given by $V_{DS} = -15$ volts and $V_{GS} = 0.6$ volt. The following biasing-circuit elements are chosen: $E = -30$ volts, $R_2 = 13\ \text{k}\Omega$, and $R_4 = 300\ \text{k}\Omega$. Find the other element values of the biasing circuit.

8-3 An *n-p-n* transistor type 2NXX17B is to be biased in the common-emitter configuration with $I_B = 0.6$ ma and $V_{CE} = 5$ volts. The *n-p-n* transistor is known to have a negative temperature coefficient. Use the biasing-circuit configuration 2 in Table 8-1 and assume that $E = 15$ volts and $R_4 = 10\ \text{k}\Omega$. Find the remaining elements of the biasing network in order to minimize the possibility of thermal runaway.

8-4 A *p-n-p* transistor type 2NXX18A is to be biased in the common-emitter configuration with $I_B = -100\ \mu$a and $V_{CE} = -5$ volts. The *p-n-p* transistor has a negative temperature coefficient. Use the biasing-circuit configuration 2 in Table 8-1. Assume $E = -15$ volts and $R_4 = 50$ kΩ. Find the remaining elements of the biasing network in order to minimize the possibility of thermal runaway.

8-5 The circuit shown in Fig. P8-5 is called a common-emitter amplifier. From various practical considerations which need not be detailed here, it is desired that this circuit be operating at $V_{CE} = 5$ volts and $I_C = 34$ ma, and that the two resistors R_1 and R_4 be related by $R_1R_4/(R_1 + R_4) = 500\ \Omega$. Without using Table 8-1, find the values of E, R_1, and R_4.

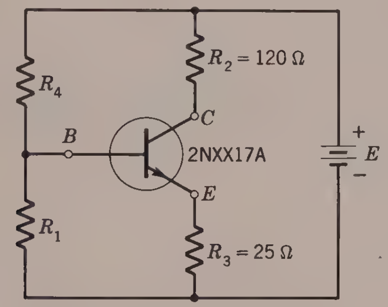
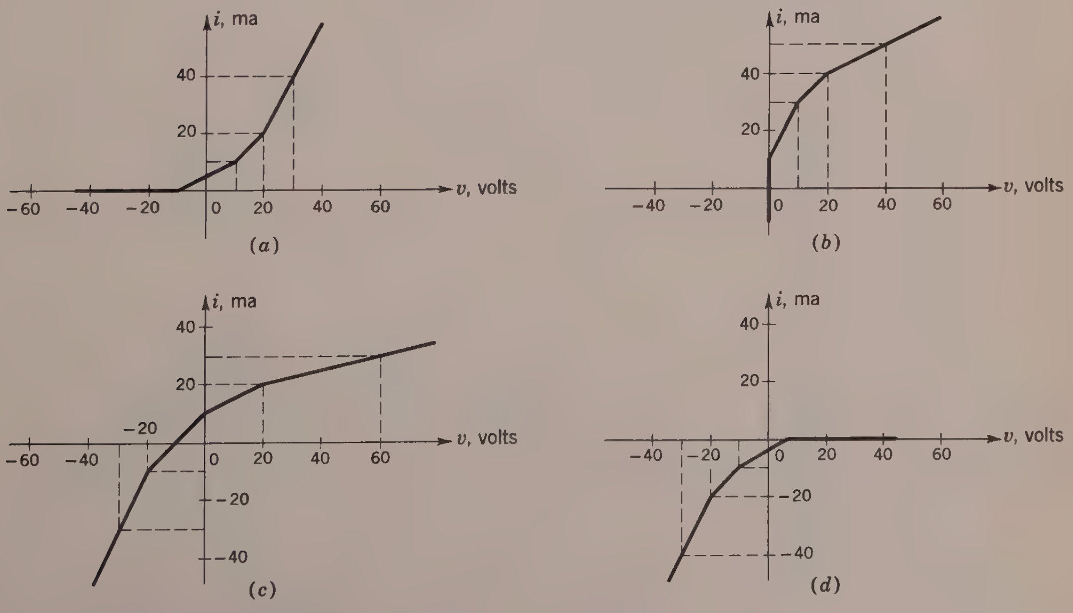


Fig. P8-5.

8-6 Synthesize the DP plots shown in Fig. P8-6*a* to *f* with concave resistors, convex resistors, and, at most, one battery.

Fig. P8-6.



(*a*) (*b*) (*c*) (*d*)

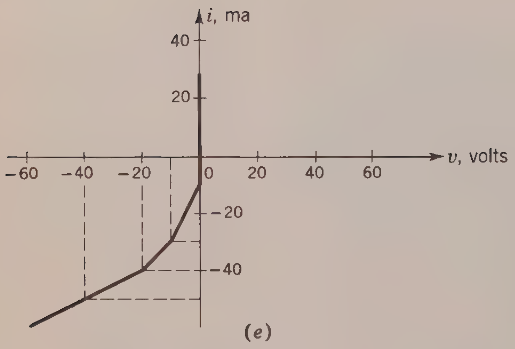


Fig. P8-6 (*Continued*).

8-7 Using only concave resistors, convex resistors, and, at most, one battery, synthesize the DP plots shown in Fig. P8-7*a* and *b*.

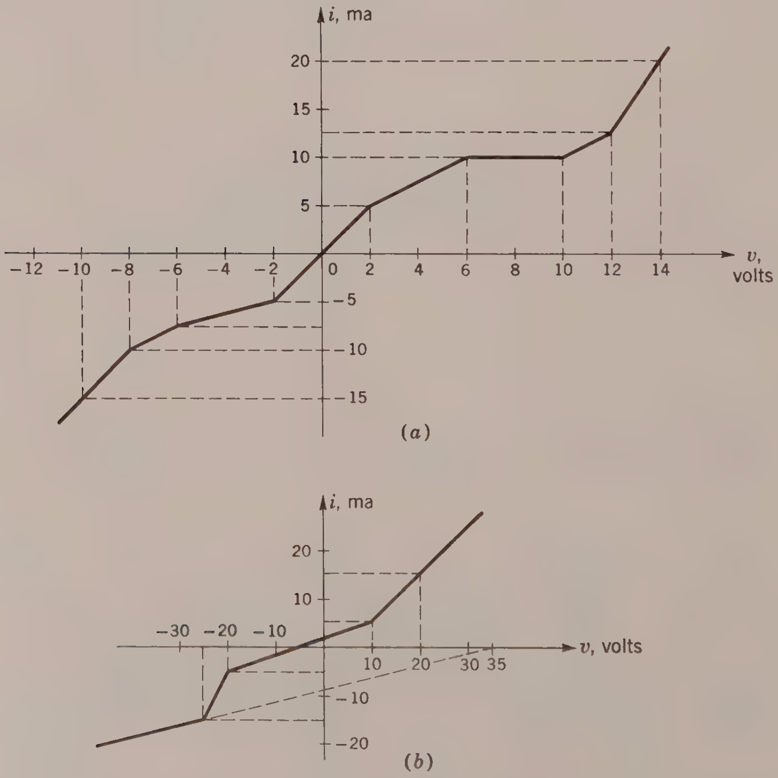


Fig. P8-7.

8-8 Synthesize the DP plots shown in Fig. P8-8*a* to *g* using only positive resistors, zener diodes, and constant-current diodes.

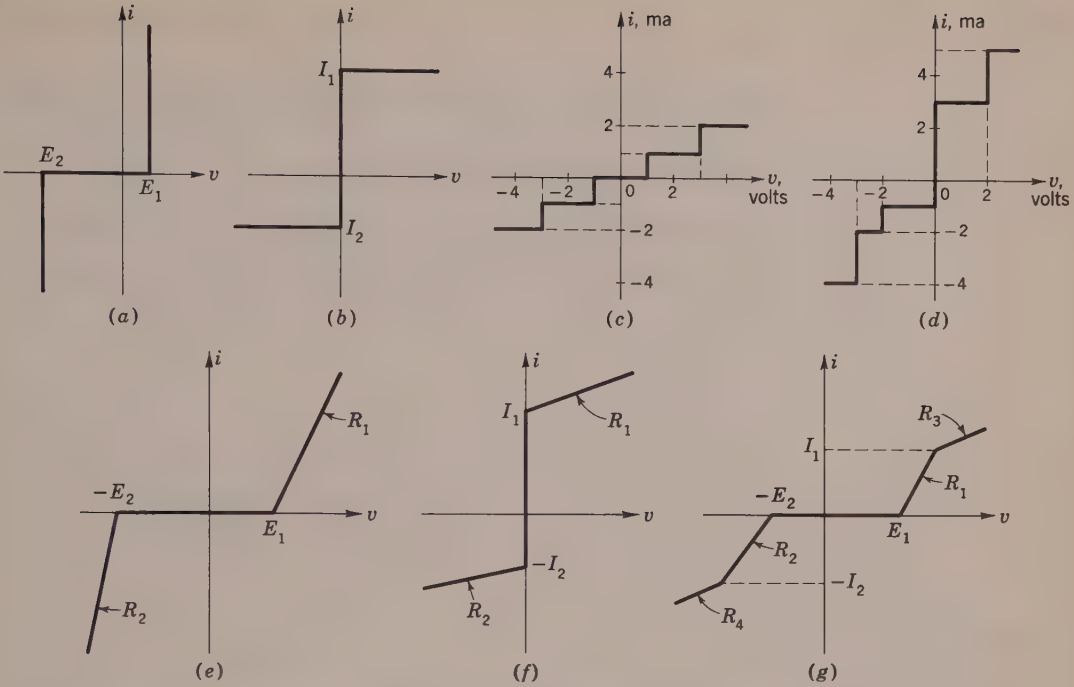
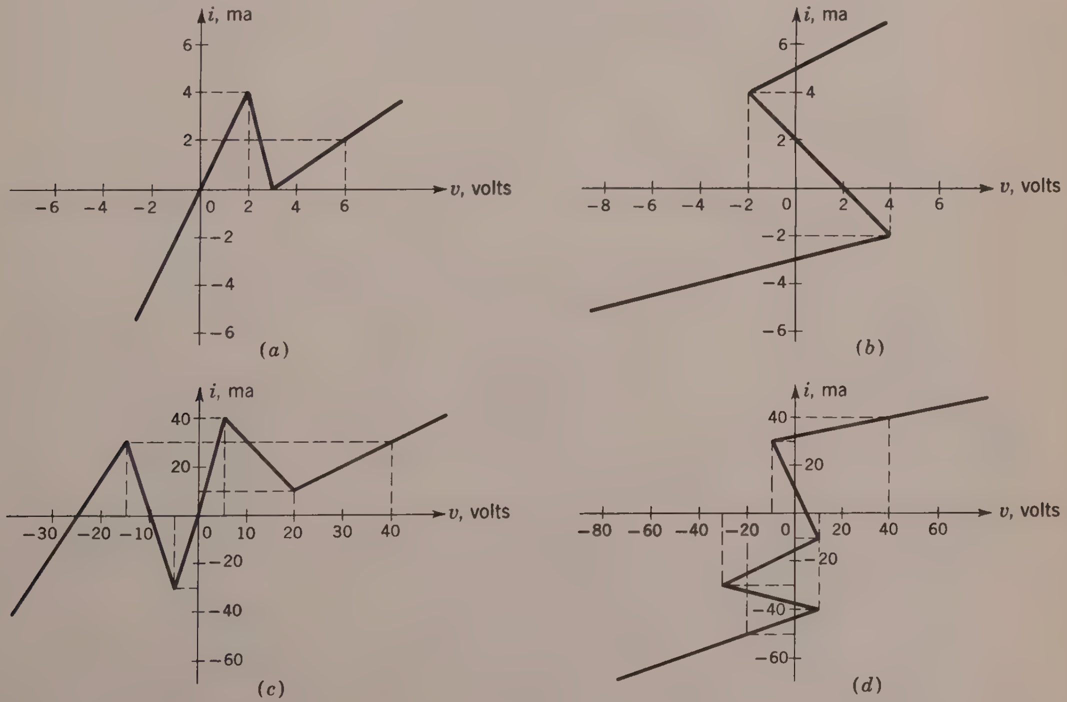


Fig. P8-8.

8-9 Using only concave resistors, convex resistors, and, at most, one battery and one negative resistor, synthesize the DP plots shown in Fig. P8-9*a* to *d*.

Fig. P8-9.

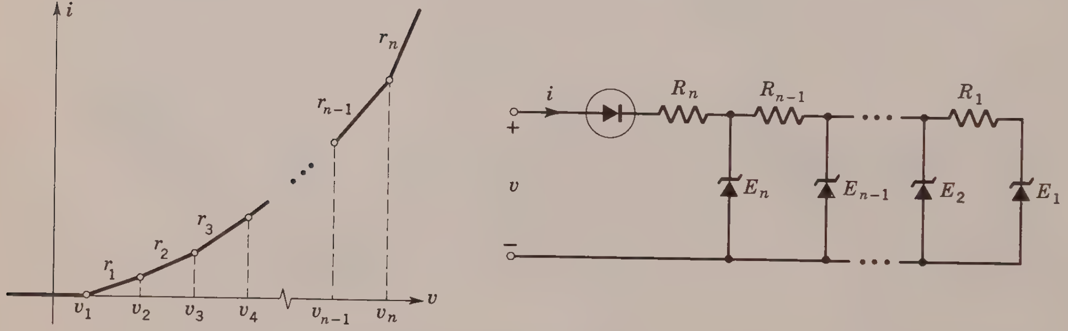


8-10 Synthesize the DP plots shown in Fig. P8-9*a* to *d* by the following methods:

(a) scalor-rotator combination
(b) scalor-reflector combination

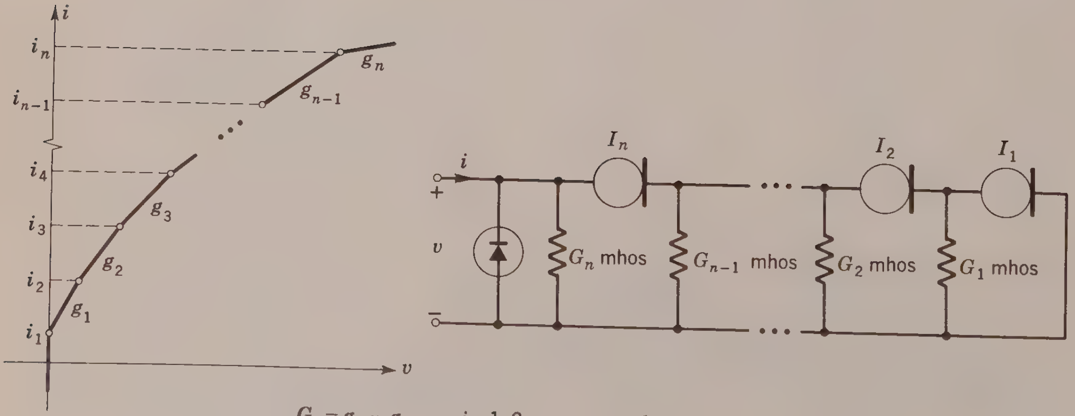
8-11 (a) Show that any concave DP plot can be realized by the zener-diode ladder network shown in Fig. P8-11*a*.
(b) Show that any convex DP plot can be realized by the constant-current diode ladder network shown in Fig. P8-11*b*.

Fig. P8-11.



$$
\begin{aligned}
R_j &= r_j - r_{j+1} && j = 1, 2, \ldots, n-1 \\
&= r_j && j = n \\
E_j &= v_j && j = 1 \\
&= E_{j-1} + R_{j-1} \sum_{k=1}^{j-1} \frac{v_{k+1} - v_k}{r_k} && j = 2, 3, \ldots, n
\end{aligned}
$$

(*a*)



$$
\begin{aligned}
G_j &= g_j - g_{j+1} && j = 1, 2, \ldots, n-1 \\
&= g_j && j = n \\
I_j &= i_j && j = 1 \\
&= I_{j-1} + G_{j-1} \sum_{k=1}^{j-1} \frac{i_{k+1} - i_k}{g_k} && j = 2, 3, \ldots, n
\end{aligned}
$$

(*b*)

8-12 Synthesize the DP plots shown in Fig. P8-6*a* and *b* by the following methods:

(a) For the concave DP plot, use the nonlinear ladder network shown in Fig. 8-10*b* of the text.

(b) Repeat (*a*) using the nonlinear ladder network shown in Fig. P8-11*a*.

(c) For the convex DP plot, use the nonlinear ladder network shown in Fig. 8-10*d* of the text.

(d) Repeat (*c*) using the nonlinear ladder network shown in Fig. P8-11*b*.

8-13 Synthesize the TC plots shown in Fig. P8-13*a* to *c* by the following methods:

(a) a nonlinear voltage divider with $R_1 = 500\ \Omega$

(b) a nonlinear voltage divider with $R_2 = 500\ \Omega$

(c) an operational amplifier with a linear feedback resistor

(d) an operational amplifier with a nonlinear feedback resistor

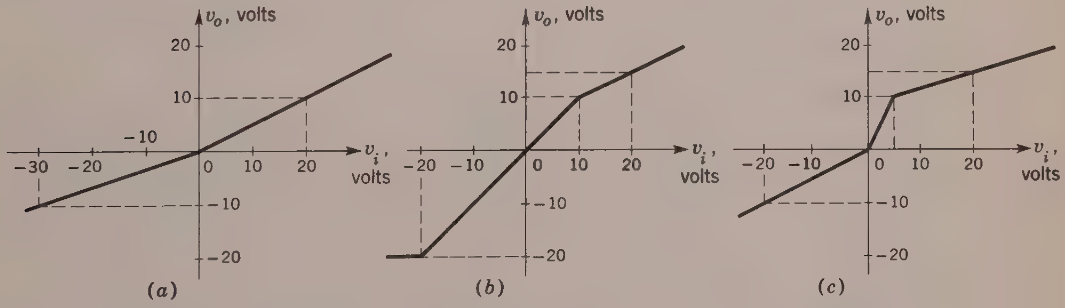


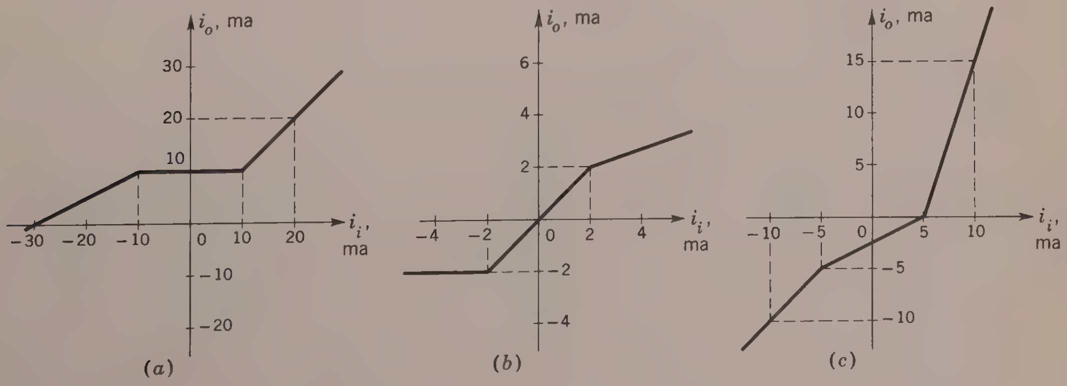
Fig. P8-13.

8-14 Synthesize the TC plots shown in Fig. P8-14*a* to *c* by the following methods:

(a) a nonlinear current-divider with $R_1 = 1{,}000\ \Omega$

(b) a nonlinear current-divider with $R_2 = 1{,}000\ \Omega$

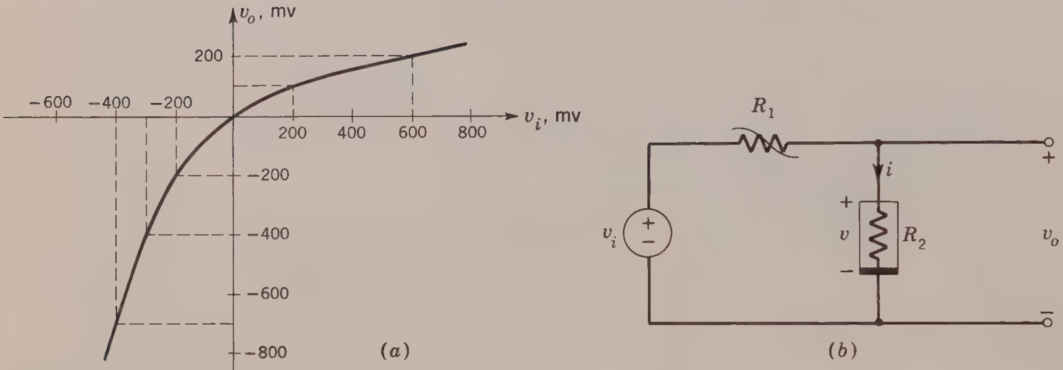
Fig. P8-14.



8-15 The TC plot shown in Fig. P8-15*a* is required to compensate for the undesirable nonlinearity of a pressure transducer. The nonlinear voltage-divider network shown in Fig. P8-15*b* is proposed for this purpose. The resistor R_1 is a varistor type 1NXX5.

(a) Find the DP plot of R_2.

(b) Synthesize R_2 using only concave resistors, convex resistors, and, at most, one negative resistor.



(*a*) (*b*)

Fig. P8-15.

8-16 When a load resistance R_L is connected across a nonlinear voltage divider as shown in Fig. P8-16*a*, the original TC plot of the network will be altered. In order to preserve the original TC plot, we can consider R_L either as a component of R_2 as shown in Fig. P8-16*b* or as a part of the resistor-source combination as shown in Fig. P8-16*c*.

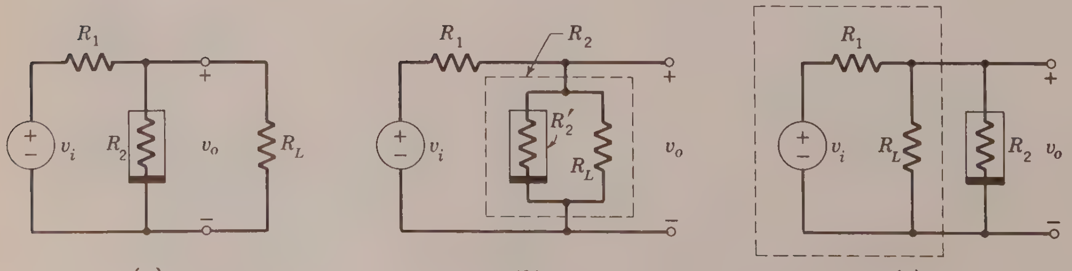
(a) Find the procedure for determining R_2' in Fig. P8-16*b*.

(b) Find the procedure for determining R_2 in Fig. P8-16*c*.

(c) Assuming that $R_L = 500\ \Omega$ and $R_1 = 500\ \Omega$, synthesize the TC plot shown in Fig. P8-13*a* with the network shown in Fig. P8-16*b*.

(d) Assuming that $R_L = 500\ \Omega$ and $R_1 = 500\ \Omega$, synthesize the TC plot shown in Fig. P8-13*c* with the network shown in Fig. P8-16*c*.

Fig. P8-16.



(*a*) (*b*) (*c*)

8-17 Show that if a TC plot Γ contains one or more segments with slopes $|m_j| > 1$, it can always be reduced to a new TC plot Γ' with segment slopes $|m_j| < 1$. This is obtained by cascading the network realizing Γ' with a voltage-controlled voltage source $v_o = kv_o'$, as shown in Fig. P8-17. If M is the maximum absolute value of the slopes of Γ, find the minimum value of k in order that the slopes of segments of Γ' be less than unity.

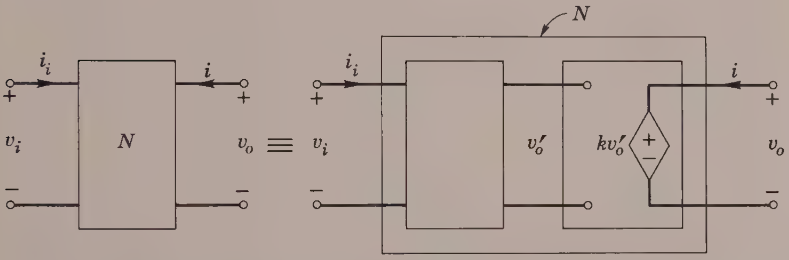
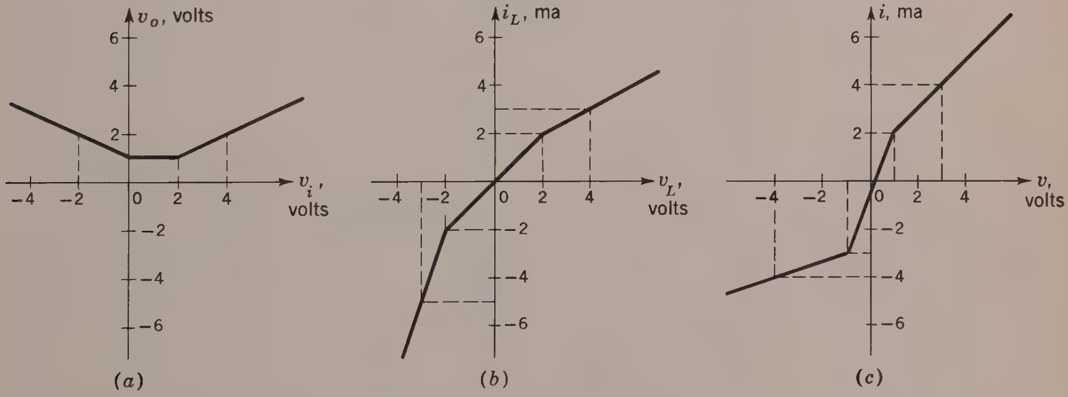


Fig. P8-17.

8-18 Using only concave and convex resistors, synthesize the jointly prescribed TC plot, load v_L-i_L curve, and DP plot shown in Fig. P8-18*a* to *c*, respectively.

Fig. P8-18.



9 SYNTHESIS OF DC-RESISTIVE NONLINEAR FUNCTIONAL NETWORKS: THE BLACK-BOX APPROACH

9-1 BASIC PHILOSOPHY

Although we have not yet studied a completely general method for analyzing arbitrary resistive nonlinear networks, we have nevertheless learned enough theory to enable us to look at some interesting practical problems. We have found it convenient and instructive to classify these practical problems into two main categories, the dc-resistive functional networks and the ac-resistive functional networks. Accordingly, the present chapter deals with the first category, and the next chapter deals with the second category.

Although no essentially new concepts will be introduced in these chapters, many new insights concerning the application of the preceding material to physical problems will be given. We hope that the practical applications considered in these chapters will provide enough motivation for the student to undertake a serious study of the general theory of resistive nonlinear networks given in subsequent chapters.

We have devoted a considerable portion of the book to the operating-point concept, the DP plot concept, and the TC plot concept. Our ultimate objective, however, is to design a network to perform a given task. We shall call this network a *functional network,* in contrast with those hypothetical networks which serve primarily as illustrations or as exercises.

Traditionally, the subject of functional networks is taught by the "cookbook approach." This usually consists of a collection of various types of functional circuits and methods for analyzing them. In this approach, the starting point is the network in question. This is followed by a brief description of the function performed by the network. An approximate analysis is then given, which usually leads to some design criteria. This procedure is applied to as many functional networks as possible during the period of time allotted. This method essentially requires that one memorize the recipes associated with a list of functional networks.

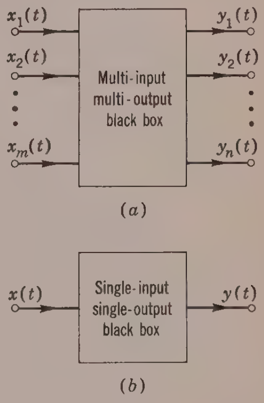
There are at least three very serious defects in the above approach. First, since the network concerned is usually given without any indication of the logical steps which led to its discovery, the procedure is not clear, and the reader is deprived of the sequence of thought required in any creative process. Second, since the method used to analyze a given network is usually tailored to suit only that particular network, it is often of a specialized nature. It is, therefore, impractical to try to fit each new network into the same mold. Third, since new networks are being discovered at an amazing rate, no one has enough time or energy to study all existing functional networks in this manner. In view of the above shortcomings, we are forced to abandon the traditional approach in favor of our black-box approach.

In the black-box approach, our starting point is always a statement of the desired task to be performed by a network. The next step is to translate this statement, if possible, into a set of network specifications. For example, the given task may consist of finding a network whose output voltage is equal to the square of the input voltage. In this case, it is easy to find an equivalent specification; i.e., synthesize the TC plot $v_o = v_i^2$. Of course, it is not always easy, or even possible, to find an equivalent specification in the form of a TC plot or a DP plot. For example, the given task may consist of finding a network whose output voltage is the derivative of the input voltage. In this case, it is clear that the specification cannot be expressed in terms of a TC plot or a DP plot. However, the specification can be expressed as $v_o(t) = dv_i(t)/dt$. Once the specification is clearly formulated, the last step consists of the synthesis of a black box that meets the specifications.

Conceptually, it is convenient to represent the task to be performed by the multi-input, multi-output black box shown in Fig. 9-1*a*. In this figure we have shown symbolically a set of input signals $x_1(t), x_2(t), \ldots, x_m(t)$ and an associated set of desired output signals $y_1(t), y_2(t), \ldots, y_n(t)$, where $x_j(t)$ or $y_j(t)$ may represent a current waveform or a voltage waveform. The black box can, therefore, be interpreted as a machine, or more precisely, a "signal processor" which processes the input signals and transforms them into the given set of output signals.

In this chapter, we shall consider mostly the simpler case where there is only one input signal $x(t)$ and only one output sig-

Fig. 9-1. The black-box representation of a multi-input, multi-output and a single-input, single-output signal processor.


$x_1(t)$
$x_2(t)$
$x_m(t)$
Multi-input multi-output black box
$y_1(t)$
$y_2(t)$
$y_n(t)$
(a)
$x(t)$
Single-input single-output black box
$y(t)$
(b)

nal $y(t)$, as symbolized by the single-input, single-output black box shown in Fig. 9-1*b*. We shall further restrict ourselves to those specifications which can be realized by dc-resistive networks, leaving those that require ac-resistive networks to the next chapter.

At this point, one may raise the objection that in view of the presence of parasitic elements, such as lead inductances and stray capacitances, any physical network is by nature a dynamic network. Why then do we study resistive functional networks? This question is best answered by stating the fact that while all physical networks are dynamic networks, some of these actually behave more like resistive networks, except possibly during brief intervals of time or during situations where the frequency of the signals is exceedingly high. We shall refer to this class of networks as *quasi-resistive networks,* in contrast with those networks which are truly dynamic networks. Since the operation of a quasi-resistive network does not depend primarily on the presence of parasitic elements, this class of networks is usually designed first as a resistive network. The effects due to the parasitic elements are then analyzed, and attempts are usually made to compensate for such undesirable effects. This is often accomplished in a number of ways, such as (1) make a careful layout of the wiring to minimize the parasitic capacitance between wires and the parasitic inductance in the leads, (2) choose high-quality components, that is, those having *v-i* curves which are valid up to some reasonably high frequencies, and (3) add small compensating inductances and capacitances at the right places. For convenience, we shall refer to the uncompensated network as the "skeleton network" and the compensation techniques as the "frostings." In line with our objective to emphasize basic concepts rather than detailed design considerations, we are interested only in the realization of a skeleton network to perform a given task. We shall leave the frostings to a more design-oriented follow-up course.

In the following sections, we shall consider various classes of common dc-resistive functional networks. Our basic philosophy is always to express the specifications in terms of one or more DP plots or TC plots or a combination of both.

9-2 FUNCTIONAL BLACK BOX FOR WAVEFORM GENERATION

There are many applications which require a specific signal waveform other than those commonly available from signal generators. For example, in automatic control systems, a control signal is usually required to achieve an optimal operation. The signal required to control the rocket motor of a spacecraft during soft-landing is a fairly complex waveform. The signals generated by many biological systems such as the brain are usually difficult to measure, and hence it is desirable that these signals be simulated by electronic means. We may classify the signals to be generated into periodic waveforms and aperiodic waveforms.

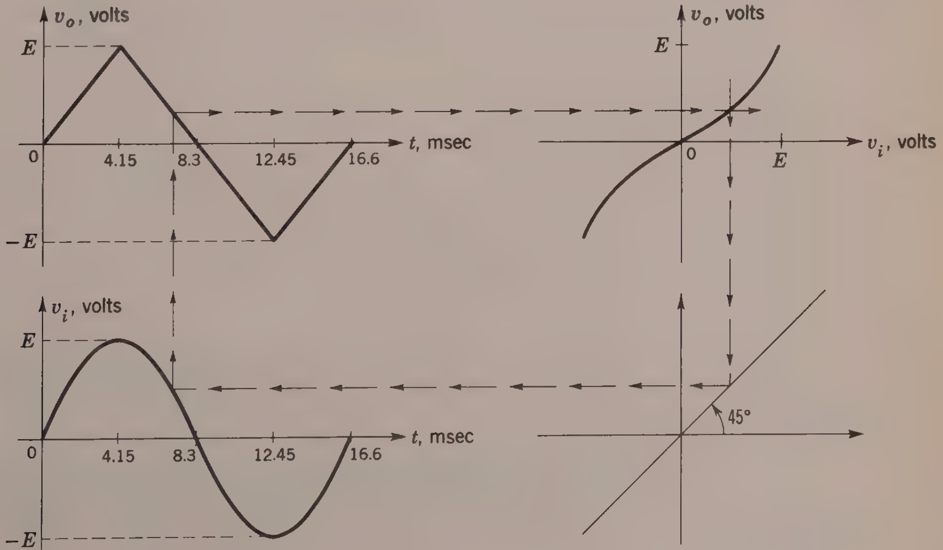
9-2-1 GENERATION OF PERIODIC WAVEFORMS

One of the simplest methods for generating a periodic waveform $v(t)$ is to transform a sinusoidal waveform into $v(t)$. This particular waveform is chosen because it can be easily generated in practice. The desired waveform $v(t)$ is obtained by applying the sinusoidal waveform across the input of a nonlinear network with an appropriate TC plot. This TC plot is easily obtained by a *graphical elimination* procedure between $v_i(t) = E \sin \omega t$ and $v_o(t)$.

EXAMPLE

A 60-Hz triangular waveform of amplitude E is required in a certain biological simulation study. To find the TC plot for generating this waveform, we choose a sinusoidal signal of *the same frequency and amplitude.* The graphical elimination procedure is applied to these two signals as shown in Fig. 9-2. The resultant TC

Fig. 9-2. The graphical elimination procedure for determining the TC plot of a sine-wave-to-triangular-wave generator.



plot shown in the upper right-hand corner of this figure can be synthesized by the methods described in Chap. 8.

The above TC plot turns out to be monotonic and hence can be easily synthesized. It is quite common, however, for the TC plot to come out nonmonotonic or even multivalued. In this case, a different input signal may lead to a more practical network realization.

9-2-2 GENERATION OF APERIODIC WAVEFORMS

If the desired waveform $v_o(t)$ is not periodic, it must be defined over some prescribed time interval $t_1 \leq t \leq t_2$. The waveform for $t \geq t_2$ is of no interest because, in practice, this signal is usually disconnected at some time t_2 from the circuit of interest. Hence, we may still use a sinusoidal input signal, provided its period T is greater than or equal to the time interval of interest, that is, $T \geq (t_2 - t_1)$.

If the above approach leads to a multivalued TC plot which is difficult to realize, we may choose another input signal waveform. A more convenient signal to choose is

$$v_i = kt \qquad t_1 \leq t \leq t_2 \tag{9-1}$$

In this case, if $v_o = f(t)$ is the desired waveform, then solving for t in Eq. (9-1) and substituting it in $f(t)$, we obtain the desired TC plot

$$v_o = f\left(\frac{v_i}{k}\right) \tag{9-2}$$

This equation guarantees that the TC plot is always a single-valued function and can, therefore, be realized by the methods described in Chap. 8.

Exercise 1: Find a practical method for realizing the TC plot shown in Fig. 9-2 and synthesize a network for this. Explain what practical considerations have been taken.

Exercise 2: Find an appropriate periodic signal waveform similar to Eq. (9-1) so that when it is applied across a network having the TC plot described by Eq. (9-2), it will generate a prescribed aperiodic waveform $v_o(t) = f(t)$ over the time interval $0 \leq t \leq T$.

9-3 FUNCTIONAL BLACK BOX FOR WAVESHAPING

The networks discussed in the preceding section require a *fixed* input signal to obtain a prescribed output signal. There is another important class of practical applications for which the input may be any signal waveform $v_i(t)$, and the output waveform $v_o(t)$ is related to $v_i(t)$ in accordance with some *specified operations.* Since the input waveforms are being *shaped* to produce a desired output waveform, we shall refer to this class of networks as *waveshaping networks.* We have already encountered many waveshaping networks in Chap. 6. In particular, any nonlinear network which performs an algebraic or transcendental operation is a waveshaping network. For example, the networks described by the TC plots $v_o = a_0 + a_1 v_i + a_2 v_i^2 + \cdots + a_n v_i^n$, $v_o = k \log v_i$, $v_o = \sqrt{|v_i|}$, $v_o = \sin v_i$, etc., are all waveshaping networks. We shall now consider some common practical examples.

9-3-1 CLIPPING NETWORKS

Given an arbitrary input signal $v_i(t)$ such as the waveform shown in Fig. 9-3*a*, it is sometimes desirable to transform this signal into an output waveform $v_o(t)$ in accordance with one of the following three sets of specifications shown in Eqs. (9-3), (9-4), and (9-5).

Case 1

$$v_o(t) = \begin{cases} v_i(t) & \text{whenever } v_i(t) \geq E_1 \\ E_1 & \text{whenever } v_i(t) < E_1 \end{cases} \tag{9-3}$$

In this case, we can imagine $v_o(t)$ as obtained by clipping off that part of the input waveform below the line $v_i = E_1$ as shown in Fig. 9-3*b*. Any network which accomplishes this task is called a *base clipper.*

Case 2

$$v_o(t) = \begin{cases} v_i(t) & \text{whenever } v_i(t) \leq E_2 \\ E_2 & \text{whenever } v_i(t) > E_2 \end{cases} \tag{9-4}$$

In this case, we can imagine $v_o(t)$ as obtained by clipping off that part of the input waveform above the line $v_i = E_2$ as shown in Fig. 9-3*c*. Any network which accomplishes this task is called a *peak clipper.*

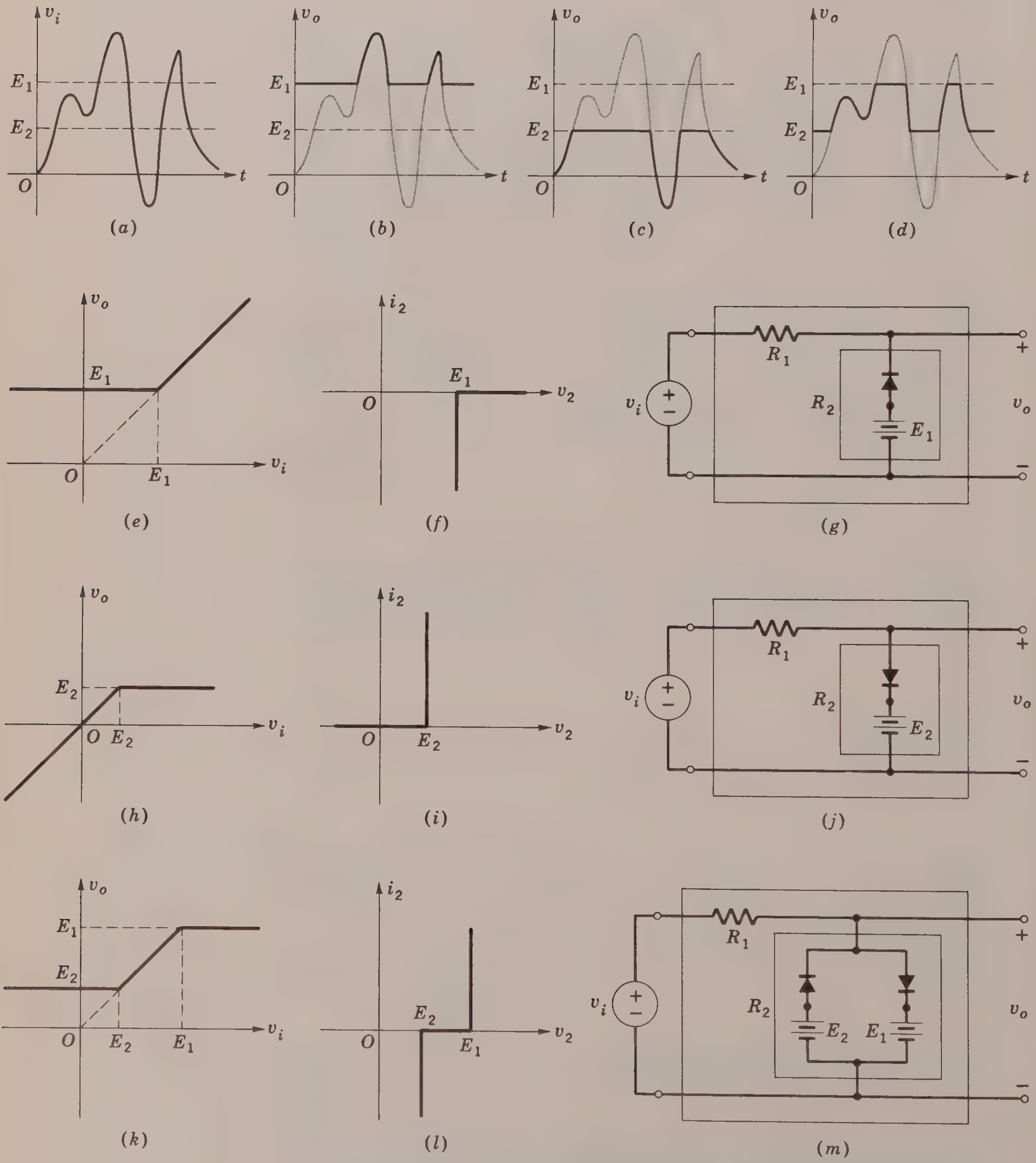


Fig. 9-3. Classification of various types of clipping networks and their realization.

Case 3

$$v_o(t) = \begin{cases} v_i(t) & \text{whenever } E_2 \leq v_i(t) \leq E_1 \\ E_1 & \text{whenever } v_i(t) > E_1 \\ E_2 & \text{whenever } v_i(t) < E_2 \end{cases} \tag{9-5}$$

In this case, $v_o(t)$ is obtained by clipping off that part of the input waveform below the line $v_i = E_2$ and above the line $v_i = E_1$ as shown in Fig. 9-3*d*. Any network which accomplishes this task is called a *base-and-peak clipper*.

Our next objective is to find the TC plot, if it exists, associated with each of the clippers defined above. An examination of the waveform in Fig. 9-3*b* and the operation defined by Eq. 9-3 leads to the TC plot shown in Fig. 9-3*e*. Since the slope of this plot is not greater than unity, we may choose the simplest realization, a nonlinear voltage divider described in Chap. 8. By the template method, we readily obtain the DP plot for the nonlinear resistor R_2 as shown in Fig. 9-3*f*. The resulting network is shown in Fig. 9-3*g*. In practice, the ideal diode is usually replaced by a junction diode. This will alter the TC plot slightly, and the resulting waveform will suffer some distortion.

An examination of the waveform in Fig. 9-3*c* and the operation defined by Eq. (9-4) results in the TC plot shown in Fig. 9-3*h*. The DP plot of the nonlinear resistor R_2 is shown in Fig. 9-3*i*, and the complete network realization is shown in Fig. 9-3*j*.

Finally, the TC plot of the base-and-peak clipper as obtained from Fig. 9-3*d* and Eq. (9-5) is shown in Fig. 9-3*k*. The DP plot of the nonlinear resistor R_2 is shown in Fig. 9-3*l*. This DP plot can be realized by connecting the two nonlinear resistors similar to R_2 of Fig. 9-3*g* and *j* in parallel, as shown in Fig. 9-3*m*. If $E_1 > 0$ and $E_2 < 0$, as is usually the case, the corresponding DP plot can be realized by two back-to-back zener diodes in series (see Fig. 7-15*d*), thereby eliminating the two batteries.

The above clipping networks are widely used in applications wherein some portion of an input signal is undesirable for subsequent operations. For example, the incoming pulse signal of many communication systems are usually contaminated with noise[1] (superimposed on top of the signal). This noise can be removed by passing the signal through a peak clipper. Another example of the use of clipping networks is in the synthesis of square waves. For example, the vertical pulse waveform shown in the upper right-hand corner of Fig. 6-37*d* can be transformed into a square wave by clipping off both the top and the bottom portion of this

[1] Any unwanted signal is called a *noise*.

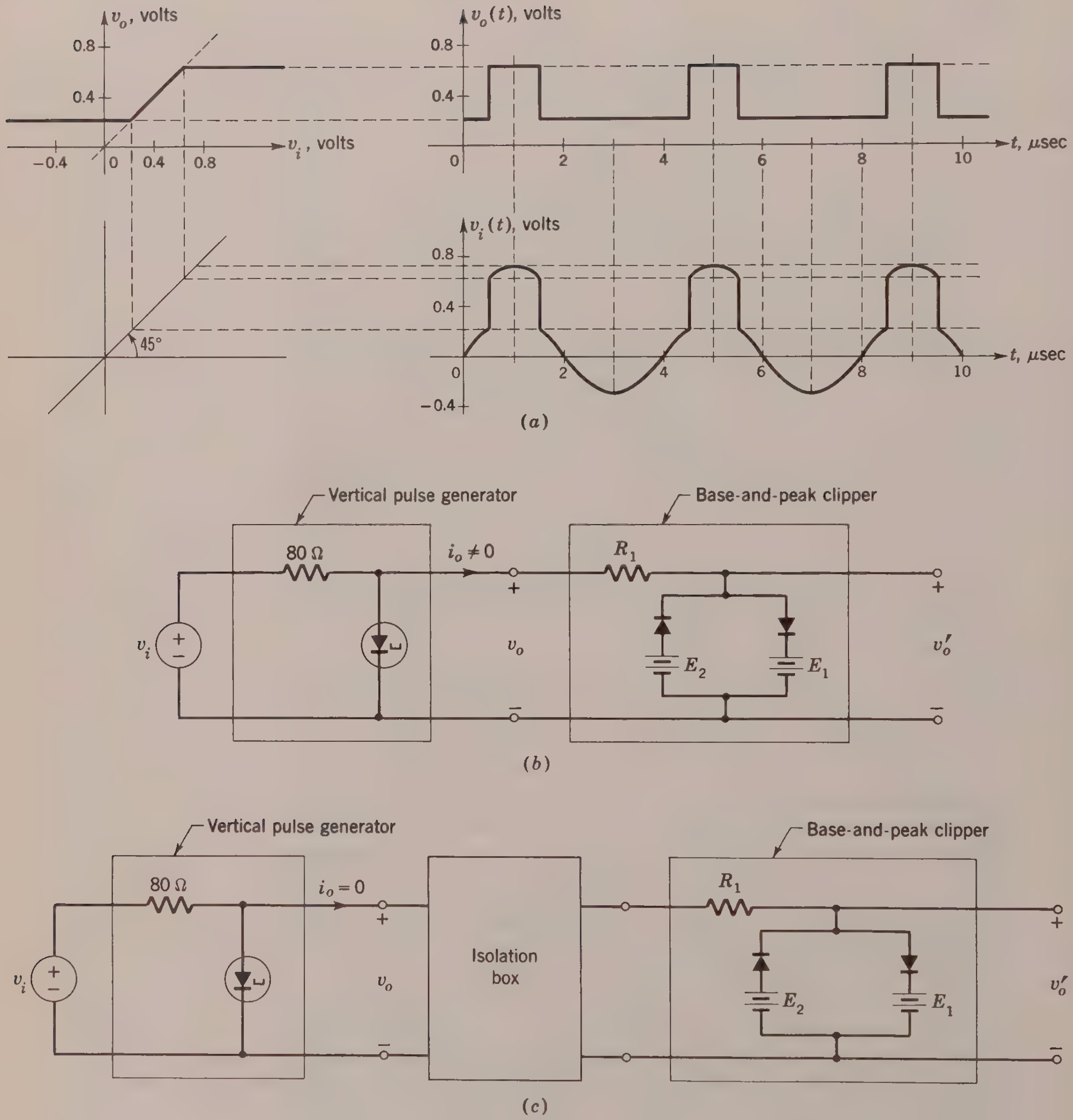


Fig. 9-4. A practical application of the base-and-peak clipper.

waveform as shown in Fig. 9-4*a*. Our first instinct suggests that we connect the tunnel-diode pulse generator of Fig. 6-37*a* across a base-and-peak clipper as shown in Fig. 9-4*b*. This is incorrect, however, because $i_o \neq 0$ in this circuit, and hence the *loading effect* of the base-and-peak clipper will alter the original TC plot of the tunnel-diode network.

There are two practical approaches to overcome the loading effects. One approach is to choose as large a value for R_1 as possible so that the DP plot seen by the tunnel diode approximates an open circuit. This method is the most economical; however, the loading effect cannot be completely eliminated. The more general approach is to "sandwich" an *isolation box* (see Fig. 3-20*c*) between the two circuits as shown in Fig. 9-4*c*.[1]

9-3-2 ANALOG-TO-DIGITAL CONVERTERS

The digital computer has often been called the most important invention of the twentieth century. It has made automation possible in virtually every field. One of the fundamental characteristics of the digital computer is that it can handle only digits. As a consequence, if we want to use a digital computer to perform some automatic operations on a given signal $v(t)$, as in the process of guiding a rocket into a proper orbit, this signal must be given as a *stepwise* or *discrete* function of time. Since most physical signals such as $v_i(t)$ in Fig. 9-5*a* are continuous functions of time, it is necessary to convert $v_i(t)$ into an approximate *discrete* signal $\bar{v}_i(t)$ as shown in Fig. 9-5*b*. A network which accomplishes this task is sometimes called an *analog-to-digital converter.* Obviously, the smaller the stepsize ΔE, the more accurate is the representation. The TC plot for an analog-to-digital converter with stepsize ΔE is easily seen to be given by Fig. 9-5*c*. Since the slope at each jump is ∞, let us choose the operational amplifier circuit (with a linear feedback resistor $R_f = 1$ kΩ) described in Chap. 8. The problem reduces to that of realizing a DP plot (in the milliampere-volt plane) which is identical with the TC plot shown in Fig. 9-5*c*. Notice that this DP plot can be easily synthesized by a series and parallel combination of the two component DP plots shown in Fig. 9-5*d* and *e*, where $k = 1, 2, \ldots, n$. These component DP plots can be realized, respectively, by two zener diodes connected back-to-back in series and two constant-current diodes connected back-to-front in parallel. The resulting ladder network shown in Fig. 9-5*f* is easily seen to be a realization of the required DP plot. If we connect this network in place of the nonlinear resistor R in the operational amplifier circuit shown in Fig. 9-5*g*, we would obtain the desired TC plot.

Exercise 1: Obtain the DP plots for the nonlinear resistor R_2 shown in Fig. 9-3*f*, *i*, and *l*. Verify the network realizations shown in Fig. 9-3*g*, *j*, and *m*.

Exercise 2: Find the effect on the output waveform resulting from our replacing the ideal diodes in Fig. 9-3*g*, *j*, and *m* by junction diodes (such as type 1NXX2).

---

[1] In practice, the isolation box may be synthesized using a *cathode follower* vacuum-tube circuit or an *emitter follower* transistor circuit.

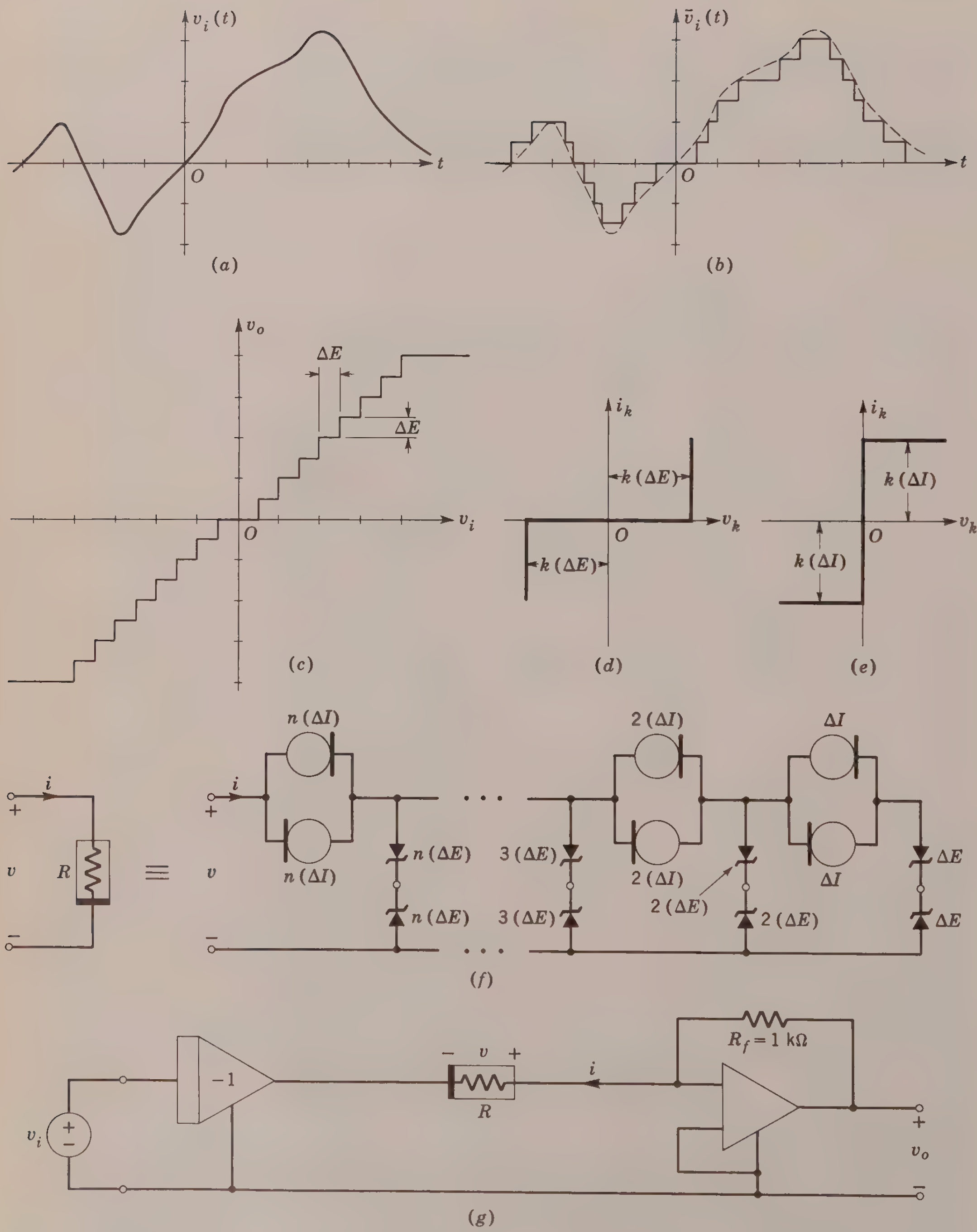


Fig. 9-5. Synthesis of an analog-to-digital converter.

Exercise 3: The duals of voltage clippers are called current clippers. (*a*) Define the three types of current clippers. (*b*) Specify the TC plots corresponding to the three types of current clippers. (*c*) Synthesize each of the three types of current clippers using only current sources, linear resistors, ideal diodes, and constant-current diodes.

Exercise 4: Show how the triangular waveform in Fig. 9-2 can be transformed into a symmetrical trapezoidal waveform. Do not neglect loading effects.

Exercise 5: Suppose the TC plot shown in Fig. 9-5*c* is to be synthesized using a nonlinear voltage divider. (*a*) If R_1 is linear, find the DP plot for R_2. (*b*) If R_2 is linear, find the DP plot for R_1.

9-4 FUNCTIONAL BLACK BOX FOR COMPENSATION

Another common problem in practice consists of designing an appropriate nonlinear network for *compensating* or *equalizing* some undesirable features of a given device, instrument, or an entire system. The basic approach to this problem consists of finding an appropriate TC plot, if it exists, for achieving this task. Once the TC plot is found, it can be synthesized by the techniques described in Chap. 8. We shall illustrate this approach by two common applications.

EXAMPLES

1. Most transducer (Fig. 9-6*a*) characteristic curves relating an output voltage v and a physical variable x are nonlinear. This nonlinearity is undesirable for several reasons. First, it is difficult to calibrate a nonlinear scale accurately because the interpolation is not linear. Second, the transducer output voltage is usually used to activate some controller mechanism whose design is greatly complicated by the nonlinearity of the transducer. A fundamental problem in transducer design, therefore, consists of designing a compensating network so that the resulting transducer characteristic curve is linear over some specified range.

   As an example, the characteristic curve of a typical *vertical-position transducer* is shown in Fig. 9-6*b*, where v is the output voltage and $x = \theta$ is the *tilt angle* from the vertical position in degrees. This type of transducer is often used aboard spacecraft, missiles, ships, etc., in order to monitor the tilt angle. To linearize this transducer, we seek a two-port compensating network connected as shown in Fig. 9-6*c*. Our problem is to determine the v_o-vs.-v TC plot of this network so that the overall transducer characteristic curve is given by $v_o = k\theta$, where k is a constant. The graphical procedure for determining the v_o-vs.-v

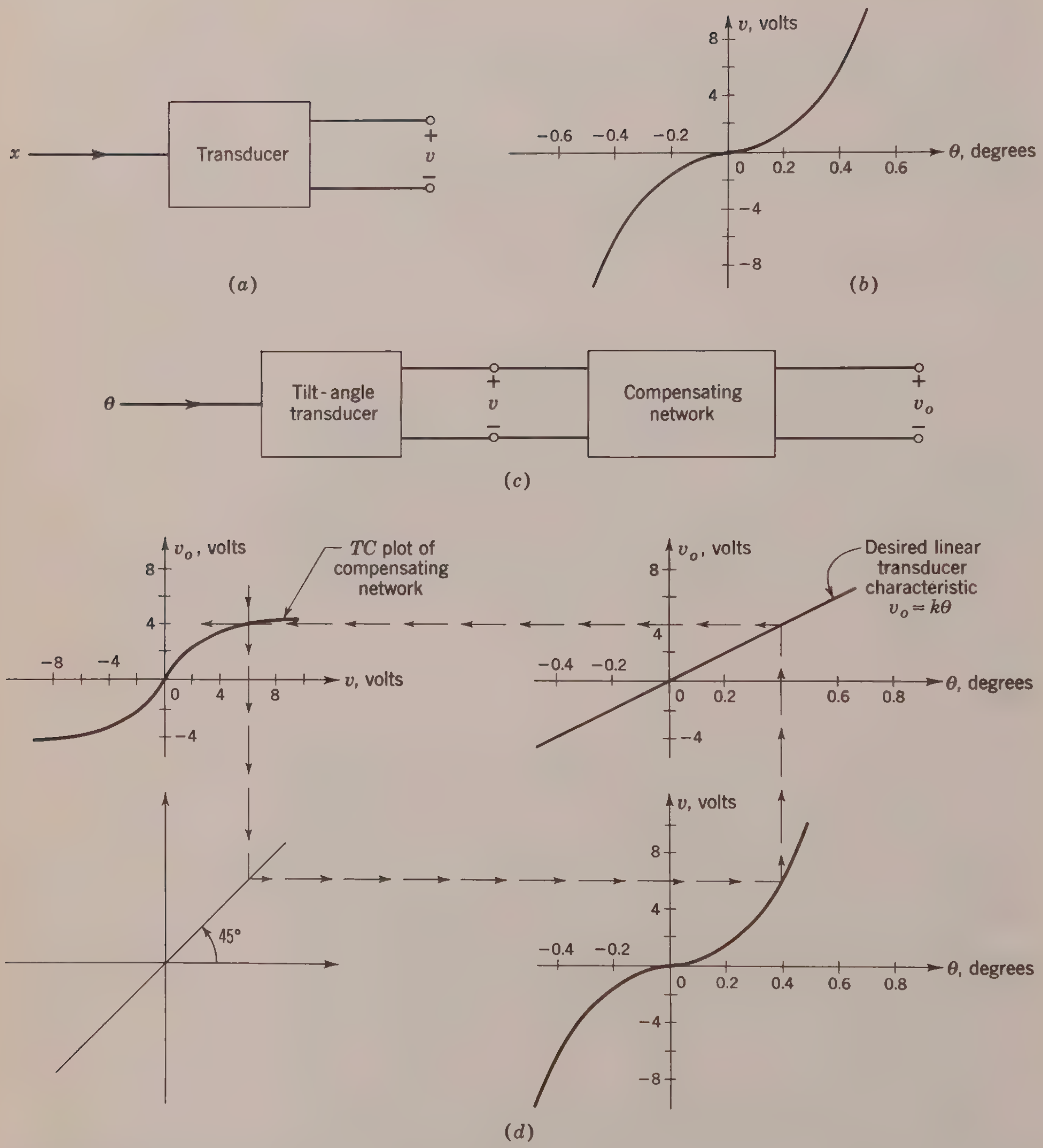


Fig. 9-6. Synthesis of a compensating network for linearizing a transducer's nonlinear characteristics.

plot corresponding to $k = 10$ is shown in Fig. 9-6*d*. The resulting TC plot in the upper left-hand corner of this figure can be easily synthesized by the techniques in Chap. 8.

2. We have already seen in Chap. 8 that the operating point of a nonlinear resistor changes with temperature. One approach for

overcoming this problem is to design a *biasing network* which also depends on temperature in an inverse manner. Another approach is to synthesize a network which produces an output voltage as a prescribed function of temperature. For example, a *varactor diode* is a nonlinear capacitor whose incremental capacitance $C(v) \equiv dq(v)/dv$ is shown in Fig. 9-7*a*. By changing the applied voltage $v = -V_Q$, the operating point Q can be moved to any point having a desired capacitance value. We have already mentioned in Chap. 1 that this property can be used to tune in radio stations by simply changing the bias voltage. The same principle can be used to change the frequency

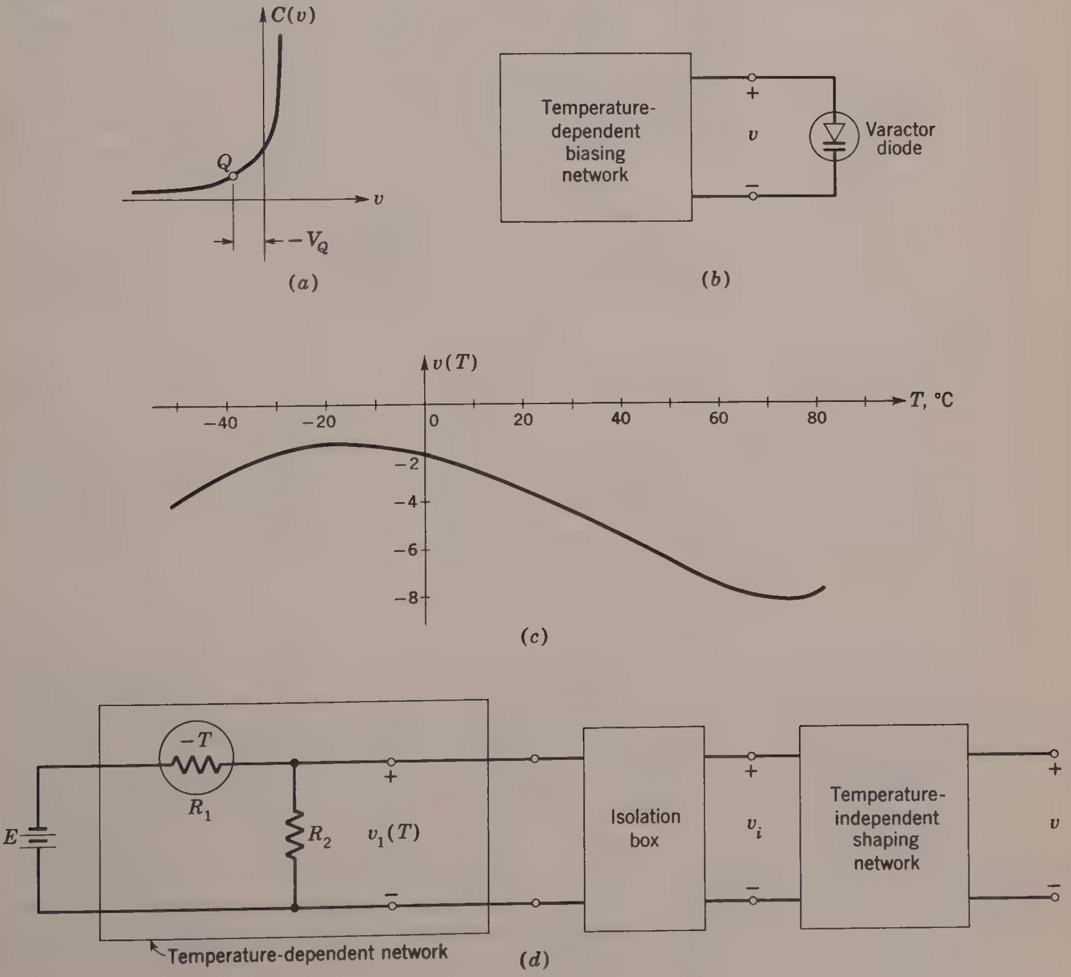


Fig. 9-7. A basic technique for producing an output voltage as a function of temperature

of oscillation of an oscillator or multivibrator.[1] Now, in practice, the frequency of any oscillator changes with temperature because of changes in the component values. However, if we measure experimentally, at each temperature, the *correction* in capacitance needed to bring back the oscillation frequency to a constant value, then we can translate this data into the form of a required biasing voltage $v(T)$ as a function of temperature. This voltage can be derived from a temperature-dependent biasing network as shown in Fig. 9-7*b*. As a specific example, the biasing voltage required to maintain the frequency of a large class of oscillators used in portable military electronic instruments is shown in Fig. 9-7*c*. In other words, if we can vary the biasing voltage in accordance with this curve, then the oscillator frequency will remain unchanged over the prescribed temperature range. One method for synthesizing the temperature-dependent biasing network is to combine thermistors in such a way as to produce the desired output voltage. This approach is satisfactory only over a small temperature range.

A more general approach is to break up the network into three parts as shown in Fig. 9-7*d*. The first part is simply a voltage-divider network where the resistor $R_1 = R_1(T)$ is *any* temperature-dependent *linear* resistor such as an NTC thermistor.[2] The output voltage across R_2 is given by:

$$v_1(T) = \frac{R_2}{R_1(T) + R_2} E \tag{9-6}$$

Since $R_1(T)$ is the resistance of an arbitrarily chosen thermistor, it is unlikely that $v_1(T)$ will fit the prescribed relationship in Fig. 9-7*c*. However, we can always find a *temperature-independent shaping network* having an appropriate $v = f(v_i)$ TC plot so that the composition

$$v = f(v_1(T)) \equiv v(T) \tag{9-7}$$

is the desired curve. Observe that what we did here was to shift the burden of realizing an appropriate *temperature-dependent* resistor $R_1(T)$ into the synthesis of a *temperature-independent* shaping network. Hence, the resistor R_1 here plays the role of a "sensing" element rather than a compensating element.

Exercise 1: In many practical problems, it is desirable to generate an output voltage v as a function of some physical variable x. Find a general procedure for achieving this task.

---

[1] It will be shown in Chaps. 16 and 17 that the frequency of oscillation of oscillators and multivibrators is a function of the capacitance in the circuit.

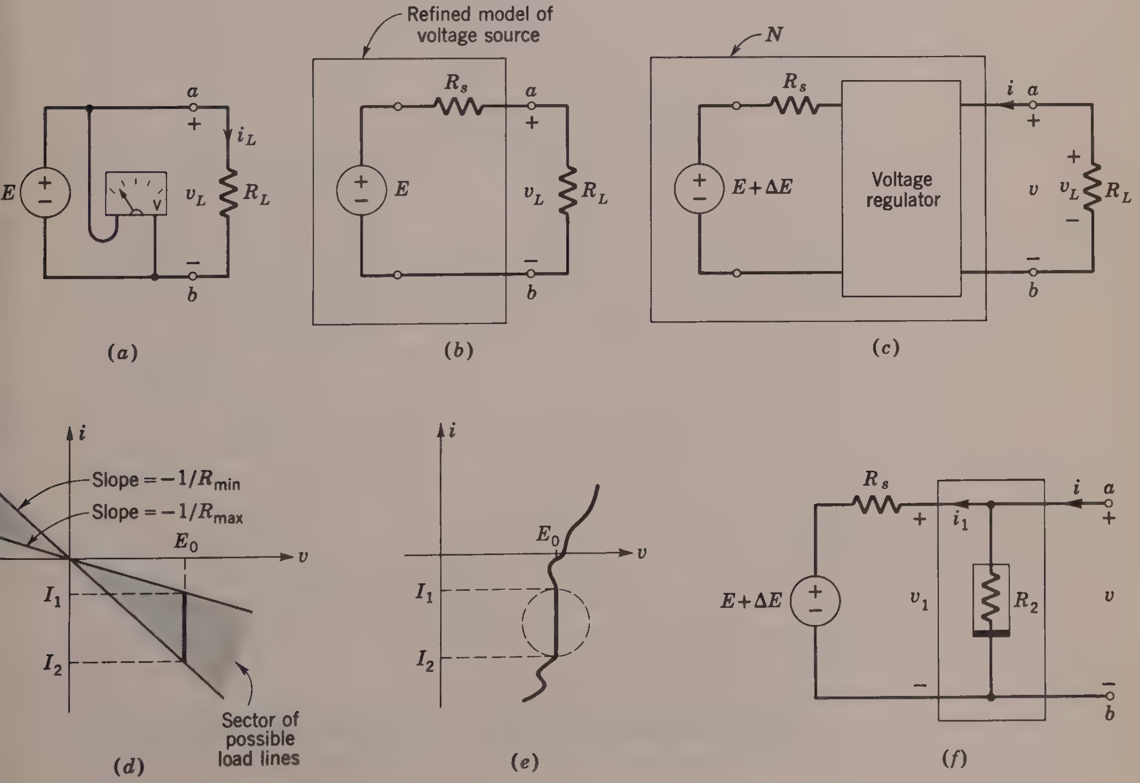
[2] When operating in its low-voltage region, the NTC thermistor can be considered as linear for most practical purposes.

Exercise 2: Show that under appropriate conditions the network to the right of the resistor R_2 (including R_2) in Fig. 9-7*d* can be replaced by a symmetric, nonlinear lattice network. Give the detailed procedure for doing this.

9-5 FUNCTIONAL BLACK BOX FOR REGULATION

If we connect a linear load resistor R_L across a *physical* voltage source, usually called a power supply, with terminal voltage E as shown in Fig. 9-8*a*, and if we measure the voltage v_L as we decrease the value of R_L, we observe that the value of v_L also decreases. This observation, of course, contradicts the definition of an ideal voltage source. To resolve this paradox, a more refined model must be chosen for the voltage source as shown in Fig. 9-8*b*, where the resistor R_s is usually called the *internal source resistance.* This results in a voltage divider where the voltage v_L is seen to decrease as R_L decreases. In addition to the effect introduced by R_s, the output voltage v_L of any physical voltage source may also vary

Fig. 9-8. The characterization of a voltage regulator in terms of a DP plot, and a possible network configuration for realizing it.



because of other external causes. For example, the internal voltage of a generator decreases whenever its prime mover slows down.

In many electronic instruments and systems, it is necessary to design a network to compensate for the above variations, so that under certain restrictions, the output voltage v_L as seen by the load resistor R_L will remain constant. To be more specific, we are given the following specifications:

1. The load resistance R_L may vary over some specified range.

$$R_{\min} < R_L < R_{\max} \tag{9-8}$$

2. The internal voltage variation ΔE may vary over some specified range.

$$\Delta E_{\min} < \Delta E < \Delta E_{\max} \tag{9-9}$$

where $\Delta E_{\min} \leq 0$ and $\Delta E_{\max} \geq 0$.

The problem is to design a two-port black box so that when it is connected between a physical voltage source and a load, as shown in Fig. 9-8*c*, the output voltage v_L remains constant at some prescribed value $v_L = E_0$, as long as the variations in R_L and ΔE fall within the above specifications. Such a two-port black box is called a *voltage regulator.*

Our first objective is to characterize the black box shown in Fig. 9-8*c*, either in terms of a TC plot or a DP plot. Since the output current i is not zero, but in fact varies as the load resistance changes, it is not meaningful to characterize the black box by a single TC plot. On the other hand, since the concept of DP plots is well defined regardless of the networks connected across the driving-point terminals, it is but natural for us to characterize the voltage regulator in terms of the DP plot seen across terminal *a-b* of the black box N, as shown in Fig. 9-8*c*. Observe that this DP plot includes the resistance R_s and the voltage source $E + \Delta E$.

Since the black box N is in parallel with R_L, the output voltage v_L is simply the intersection of the DP plot of N and the load line $v = -R_L i$, where the value of R_L may vary according to Eq. (9-8). This means that if we draw the load lines corresponding to the two extreme values $R_{\max}$ and $R_{\min}$, as shown in Fig. 9-8*d*, then any intermediate load line will necessarily fall within the sector bounded by these two load lines. In order to obtain a constant output voltage $v_L = E_0$ for all values of load resistance R_L within the specified range, it is necessary that the DP plot of N be a vertical segment inside this sector, as shown in Fig. 9-8*d*. The shape of the DP plot outside the sector is irrelevant and may assume any convenient form, such as the plot shown in Fig. 9-8*e*.

Our next objective is to find a two-port network so that when one of its ports is connected to a physical voltage source, the DP plot across the other port satisfies the above condition. Many two-port networks can be found. We shall consider only the simplest case, that is, a resistor R_2 in place of the two-port network, as shown in Fig. 9-8*f*. In order to find the DP plot of R_2, we reverse the procedure for obtaining DP plots of two resistors in parallel. This can be done by subtracting the ordinates of the pertinent portion of the DP plot of N, as shown in Fig. 9-9*a*, from the DP plot of the resistor-voltage source combination (corresponding to a typical generator voltage $E + \Delta E$), as shown in Fig. 9-9*b*. The resulting v_2-i_2 curve contains a vertical segment in the first quadrant between $i = (I_1 - I_3)$ and $i = (I_2 - I_3)$ as shown in Fig. 9-9*c*. Since the remaining portion of the DP plot is immaterial, we are free to choose any commercially available resistor which meets the above specification in the first quadrant. This immediately suggests a zener diode connected as shown in Fig. 9-9*d*. For flexibility, we have added a resistor R in series with the internal resistance R_s. To obtain the DP plot of N, we draw the v_2-i_1 curve of the resistor-voltage source combination and the v_2-i_2 curve of the zener diode as shown in Fig. 9-9*e*. Observe that since ΔE may vary according to Eq. (9-9), we have shown two v_2-i_1 curves corresponding to the two extreme cases. The resulting DP plot of N corresponding to an intermediate value $E + \Delta E$ is shown in Fig. 9-9*f*. Observe that in order for this DP plot to satisfy the requirement shown in Fig. 9-8*e*, it is necessary that $I_4 \leq I_2$. This requirement can be easily met by a suitable choice of the series resistance R. The complete voltage-regulator network is shown in Fig. 9-10*a*.

In practice, the zener diode is rated by the maximum current $I_{\max}$ that it can safely handle. Therefore, we must also satisfy the condition

$$i_2 < I_{\max} \tag{9-10}$$

From Fig. 9-10*a* we obtain

$$i_2 = i - i_1 \tag{9-11}$$

The value of i is found by determining the operating point Q of the network as shown in Fig. 9-10*b*. Let this be denoted by

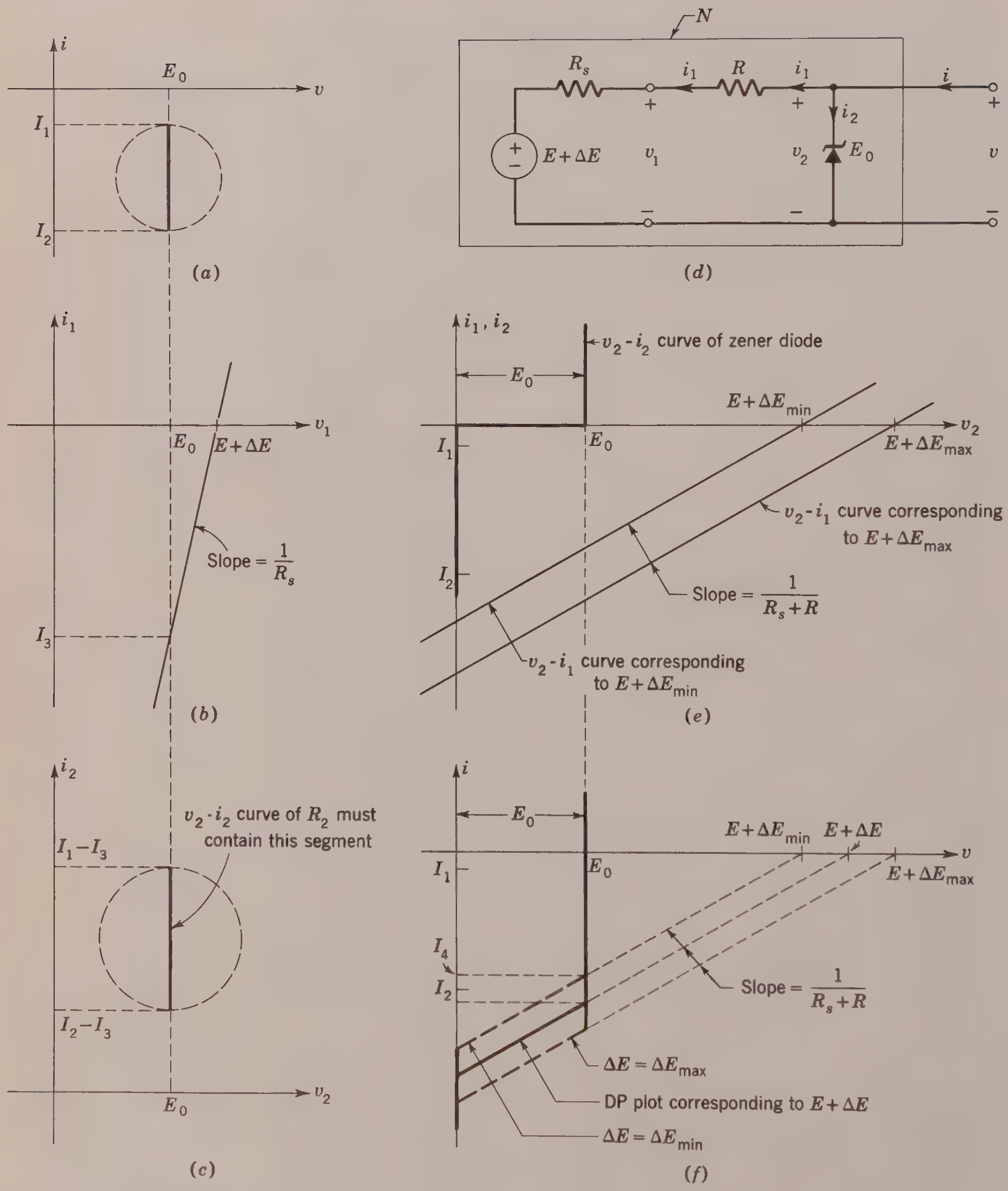


Fig. 9-9. The synthesis of a voltage regulator using one zener diode.

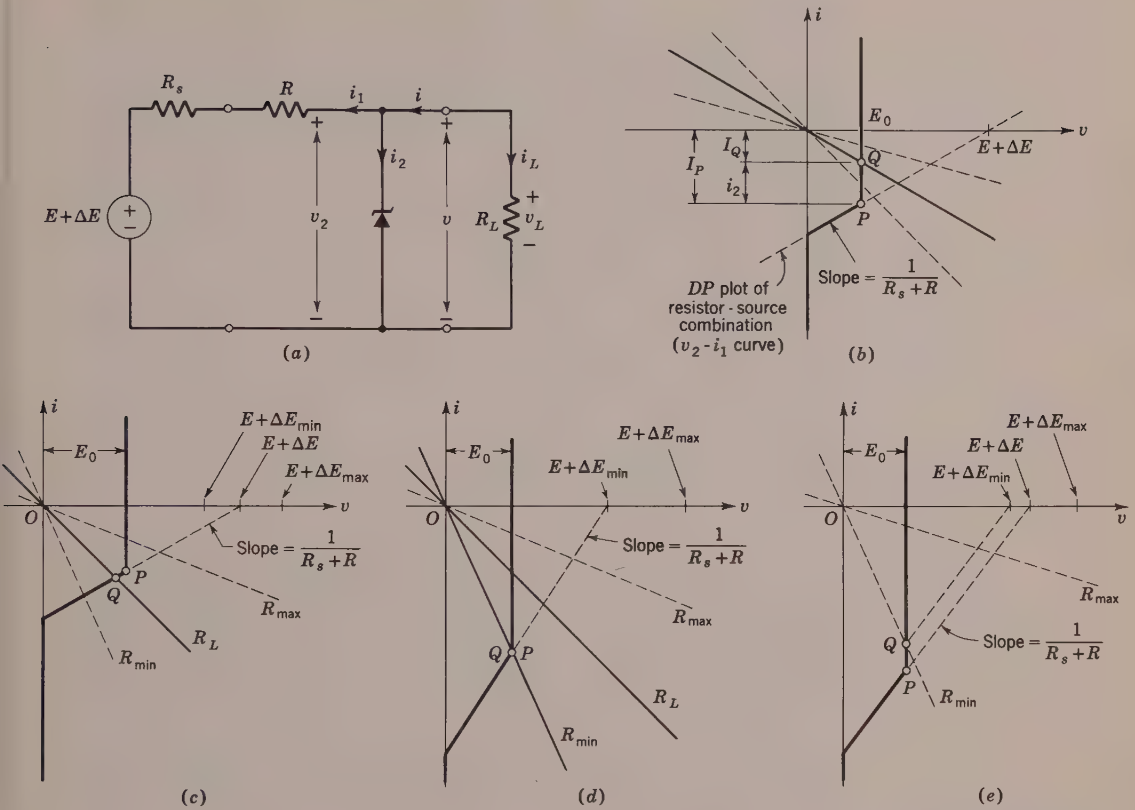

R_s
R
i_1
i
i_2
i_L
$E+\Delta E$
v_2
v
R_L
v_L
(a)
i
E_0
v
$E+\Delta E$
I_Q
I_P
i_2
Q
P
DP plot of resistor - source combination (v_2-i_1 curve)
Slope = $\frac{1}{R_s+R}$
(b)
i
E_0
$E+\Delta E_{\text{min}}$
$E+\Delta E$
$E+\Delta E_{\text{max}}$
v
O
Slope = $\frac{1}{R_s+R}$
P
Q
R_{max}
R_L
R_{min}
(c)
i
E_0
$E+\Delta E_{\text{max}}$
$E+\Delta E_{\text{min}}$
v
O
Slope = $\frac{1}{R_s+R}$
R_{max}
Q
P
R_L
R_{min}
(d)
i
E_0
$E+\Delta E_{\text{max}}$
$E+\Delta E$
$E+\Delta E_{\text{min}}$
O
v
R_{max}
Q
Slope = $\frac{1}{R_s+R}$
P
R_{min}
(e)

Fig. 9-10. The complete voltage-regulator network and the graphical procedure for determining the optimum value of the current-limiting resistor R.

$i = -I_Q$, where I_Q is the distance shown in Fig. 9-10*b*. The value of i_1 is found by substituting the output voltage $v_2 = E_0$ into the DP plot equation

$$v_2 = (R_s + R)i_1 + (E + \Delta E) \tag{9-12}$$

Observe that this DP plot corresponds to the dotted line through $v = E + \Delta E$ in Fig. 9-10*b*.

Therefore, the value of i_1 is simply the intersection P between the DP plot and the vertical line through $v = E_0$ as shown in Fig. 9-10*b*. Let us denote this current by $i_1 = -I_P$, where I_P is the distance shown in Fig. 9-10*b*.

Substituting the values of i and i_1 into Eq. (9-11), we obtain

$$i_2 = I_P - I_Q \tag{9-13}$$

This equation shows that the distance between points P and Q is equal to the magnitude of the current flowing through the zener diode. Hence, in order to satisfy Eq. (9-10), we must be sure that point P must be close to point Q.

Observe that if we decrease the value of R but keep the internal voltage source constant, the breakpoint P will move downward because the slope of the v_2-i_1 curve decreases. This will increase the distance between points Q and P, thereby increasing the current in the zener diode. The role played by the resistor R is now obvious: it limits the current drawn by the zener diode so that condition (9-10) can be satisfied.

One might be led to conclude at this point that R should be chosen as large as possible in order to prevent "burning out" the zener diode. That this is not true is readily seen by the load-line construction shown in Fig. 9-10*c*. Observe that because R is too large, the operating point Q corresponding to some load resistance R_L no longer falls on the vertical segment, and the output voltage is different from E_0.

There must, therefore, be an optimum value for R. This can be found by examining the worst case that might happen, that is, when $R_L = R_{\min}$ and $\Delta E = \Delta E_{\min}$. The corresponding load-line construction is shown in Fig. 9-10*d*. Clearly, the largest value of R that can be chosen occurs when the operating point Q corresponding to the *minimum* load $R_{\min}$ coincides with the breakpoint P. With this optimum value of R, it is seen that for any other value of generator voltage $\Delta E > \Delta E_{\min}$, such as shown in Fig. 9-10*e*, the operating point Q will always lie along the vertical segment $v = E_0$.

Although the above discussion is restricted only to a simple zener-diode voltage-regulator circuit, the same principle can be extended to analyze and synthesize other types of voltage-regulator circuits. In particular, the DP plot seen across the load resistor must necessarily contain a vertical segment as in Fig. 9-8*e*.

Exercise 1: The dual case of a voltage regulator is a current regulator. Find a simple realization.

Exercise 2: Find the effect on the voltage regulation when an actual zener diode with an almost, but not quite, vertical segment is used in place of the ideal zener diode.

9-6 SUMMARY

Basic philosophy Given a set of specifications involving the processing of signals, our first step is to translate the functions to be performed into circuit theoretic terms. In many practical cases, the specifications can be expressed in terms of a combination of one or more TC plots and DP plots. Once this equivalent specification is found, the techniques discussed in Chap. 8 can be used to realize the pertinent networks. The advantage of this approach is that any network which accomplishes the same task must be described by the same combination TC plots and DP plots. This makes it unnecessary to study each circuit as a separate entity.

Functional black box for waveform generation The network is designed to transform a *specific* input waveform into a *prescribed* output waveform.

1. A periodic waveform is usually generated by applying a sine wave of the same frequency across a network with an appropriate TC plot.
2. An aperiodic waveform can always be generated by applying a linear time function $v_i = kt$ across a network with an appropriate TC plot.

Functional black box for waveshaping The signal-processing operation is applicable to *arbitrary* input signals. For example, a clipping network is designed to clip off the lower portion (base clipper), the upper portion (peak clipper), or both the lower and upper portions (base-and-peak clipper) of an input signal.

Functional black box for compensation This class of network is used to correct or equalize some undesirable features of an existing device, circuit, or system. Two common applications are as follows:

1. To linearize the nonlinear characteristics of transducers.
2. To generate a temperature-dependent output voltage.

Functional black box for regulation This class of network is used to regulate the output signal waveforms over a wide range of load resistors. The most common example is a voltage regulator.

PROBLEMS

9-1 Given an input waveform $v_i(t)$ and an output waveform $v_o(t)$, the TC plot obtained from a graphical elimination between these signals may be a monotonic function, a single-valued function, or a multivalued function.

(a) Find some simple criterion in terms of $v_i(t)$ and $v_o(t)$ so that the TC plot is monotonic.

(b) Find some simple criterion in terms of $v_i(t)$ and $v_o(t)$ so that the TC plot is a single-valued function.

9-2 A triangular waveform of either polarity as shown in Fig. P9-2*a* and *b* is available. Design a network to transform these signals into the output waveforms shown in Fig. P9-2*c* and *d*, respectively. Such a network is called a *frequency doubler*.

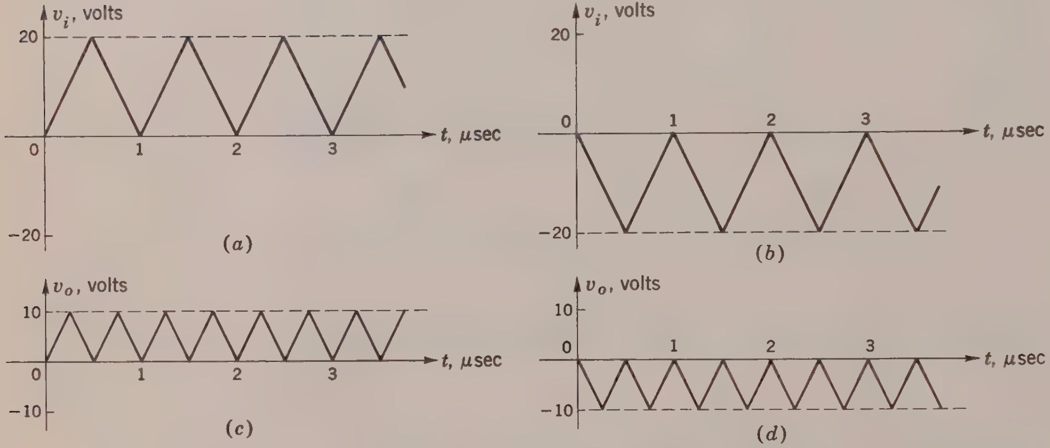


Fig. P9-2.

9-3 A third harmonic sinusoidal generator is a signal processor for converting an input voltage $v_i(t) = E \sin \omega t$ into an output voltage $v_o(t) = kE \sin 3\omega t$.

(a) Find the TC plot of a third harmonic sinusoidal generator corresponding to $k = -\frac{1}{2}$, 1, and 2.

(b) Design a third harmonic generator with $E = 10$ volts and $k = \frac{1}{2}$.

(c) Repeat (*b*) with $k = 2$.

9-4 An *n*th harmonic sinusoidal generator is a signal processor for converting an input voltage $v_i(t) = E \sin \omega t$ into an output voltage $v_o(t) = kE \sin n\omega t$.

(a) Find the nature of the TC plot if n is an odd integer. Is it a single-valued function?

(b) Repeat (*a*) for even integers.

(c) Repeat (*a*) and (*b*) for nonsinusoidal periodic signals.

9-5 An *n*th-order subharmonic sinusoidal generator is a signal processor for converting an input voltage $v_i(t) = E \sin \omega t$ into an output voltage $v_o(t) = kE \sin (\omega/n)t$.

(a) Find the nature of the TC plot if n is an odd integer. Is it a single-valued function?

(b) Repeat (*a*) for even integers.

(c) Repeat (*a*) and (*b*) for nonsinusoidal periodic signals.

9-6 In a medical experiment a dog's heart is to be stimulated by a periodic pulse of the form shown in Fig. P9-6*a*, where the quantities T, τ, and E are all variable. A signal generator capable of generating a symmetrical, positive, triangular waveform with arbitrary amplitude and frequency is available. Design a network for accomplishing this task. Use the output voltage of the signal generator as the input of this network.

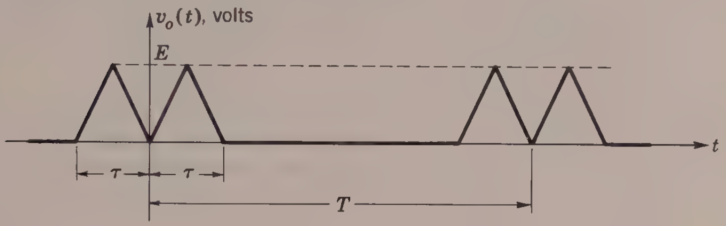


Fig. P9-6.

9-7 Find the TC plot of a signal processor for converting an input voltage $v_i(t)$ into an output voltage $v_o(t)$ according to the following rule:

$$v_o(t) = \begin{cases} v_i(t) & \text{whenever } v_i(t) \geq 3 \text{ volts} \\ 3 & \text{whenever } v_i(t) < 3 \text{ volts} \end{cases}$$

(a) Synthesize the above TC plot using an ideal diode and a battery. If $v_i(t) = 6 \sin 2\pi t$, sketch the output voltage.

(b) If the ideal diode is replaced by a junction diode type 1NXX2 and the same input signal in (*a*) is applied, find the resulting output voltage. Compare this with the desired output voltage as obtained in (*a*).

(c) In order to minimize the discrepancy between the actual output signal obtained in (*b*) and the desired output signal, the battery voltage E of the network obtained in (*a*) should be changed when the junction diode is introduced in place of the ideal diode. Find the new voltage E.

9-8 Repeat Prob. 9-7 for the case where the output signal is given by the following rule:

$$v_o(t) = \begin{cases} v_i(t) & \text{whenever } v_i(t) \geq -3 \quad \text{volts} \\ -3 & \text{whenever } v_i(t) < -3 \quad \text{volts} \end{cases}$$

9-9 Repeat Prob. 9-8 for the case where the output signal is given by the following rule:

$$v_o(t) = \begin{cases} v_i(t) & \text{whenever } v_i(t) \leq -3 \quad \text{volts} \\ -3 & \text{whenever } v_i(t) > -3 \quad \text{volts} \end{cases}$$

9-10 In a certain instrumentation problem, it is required to design a network so that the output current is the natural logarithm of the input current; that is,

$$i_o(t) = \begin{cases} \ln i_i(t) & \text{whenever } i_i(t) \geq 1 \text{ ma} \\ 0 & \text{whenever } i_i(t) < 1 \text{ ma} \end{cases}$$

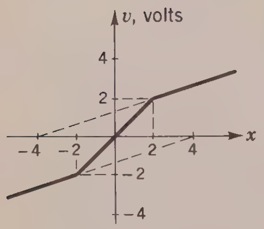


Fig. P9-11.

(a) Design a network for accomplishing this specification.
(b) Repeat (*a*) for the dual case where the variable i is replaced by the voltage v.

9-11 The v-vs.-x relationship of a certain pressure transducer for measuring eyeball pressure is shown in Fig. P9-11. Design a nonlinear compensation network so that the overall characteristic of the transducer is given by $v_o = \frac{1}{4}x$.

9-12 It is desired to design a thermistor network in a refrigerator-oven environmental chamber to give an output voltage $v_o(t)$ which is *linearly* related to the temperature inside the chamber. Such an arrangement would allow the voltage across a temperature-sensitive device under observation (inside the chamber) to be plotted directly versus temperature by an x-y recorder whose movements are controlled by the device's output voltage. As a starting point in the design, consider the themistor voltage divider circuit shown in Fig. P9-12*a*.

(a) If the thermistor is assumed to be operating on its linear region with a resistance $R_1(T)$ as shown in Fig. P9-12*b*, find the voltage $v(T)$ across the 1-kΩ resistor R_2.
(b) Find the TC plot of the compensating network so that the overall characteristic is given by $v_o = 1 + T/100$ volts.
(c) Synthesize the complete network using only concave resistors, convex resistors, a battery, and an isolation box.
(d) Synthesize the TC plot in (*b*) by a symmetric lattice network.

HINT: Replace the 1-kΩ resistor R_2 in Fig. P9-12*a* by an appropriate nonlinear lattice network with an equivalent 1-kΩ linear DP plot.

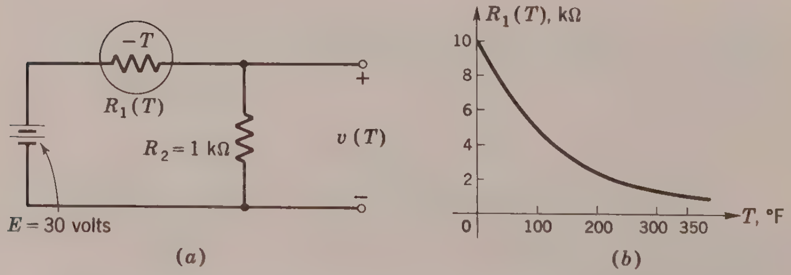


Fig. P9-12. (*a*) (*b*)

9-13 The most commonly used temperature-sensing element in the design of temperature-dependent networks is the NTC thermistor. When operating at low voltages, the thermistor can be represented by a temperature-dependent linear resistance

$$R(T) = R_0 e^{\beta(1/T - 1/T_0)}$$

where T = absolute temperature of thermistor
R_0 = resistance at temperature T_0
β = constant which depends on material of which thermistor is made

Thermistors with a wide range of R_0 and β are commercially available. It is possible to "shape" the resistance-vs.-temperature curve of a thermistor by combining it with other elements. The five most common combinations are shown in Fig. P9-13*a* to *e*.

(a) What is the effect of R_1 on the resistance-vs.-temperature curve of the composite elements in Fig. P9-13*a* and *b*?

(b) What is the effect of the choice of β and R_0 of the two thermistors $R_1(T)$ and $R_2(T)$ in Fig. P9-13*c* and *d* on the overall resistance-temperature curve?

(c) What is the effect of R_1, R_2, and R_3 on the overall resistance-temperature curve in Fig. P9-13*e*?

(d) Find the general shape of the resistance-vs.-temperature curve of the composite elements in Fig. P9-13.

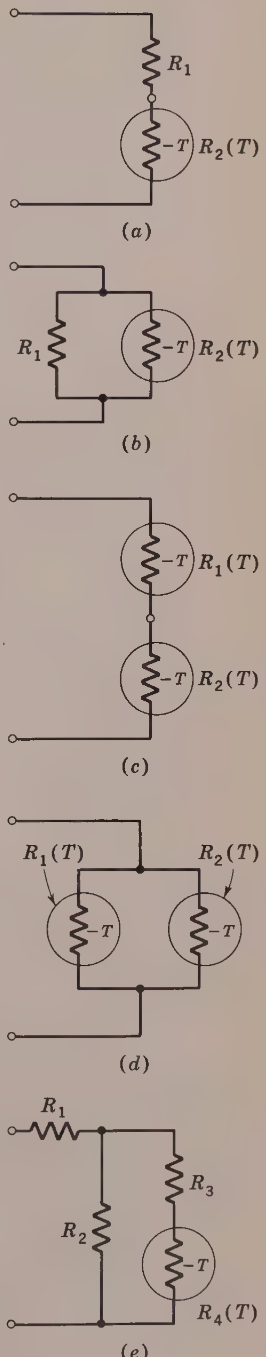


Fig. P9-13.

9-14 A certain voltage power supply can be represented by a 15-volt internal voltage source in series with a 300-Ω internal source resistance. The internal voltage source is known to have a maximum voltage variation of ±2 volts. If the load resistance to be connected to the power supply is at least 500 Ω, design a voltage regulator to maintain the output voltage constant at 6 volts. The available zener diodes have a maximum current rating of 20 ma. You may assume that the zener diodes have perfectly vertical segments.

9-15 The negative-resistance portion of a thermistor can be utilized to design voltage regulators. The basic circuit is shown in Fig. P9-15. Assume that the thermistor is maintained at a fixed temperature of 50° C.

(a) Plot the v_o-vs.-v_i TC plot.

(b) Determine the approximate range of input voltages for obtaining almost constant output voltage.

(c) Repeat (*a*) and (*b*) with $R_L = 50$ kΩ.

(d) Repeat (*a*) and (*b*) with $R_L = 1$ kΩ.

(e) Compare the amount of regulation for the above three different values of R_L.

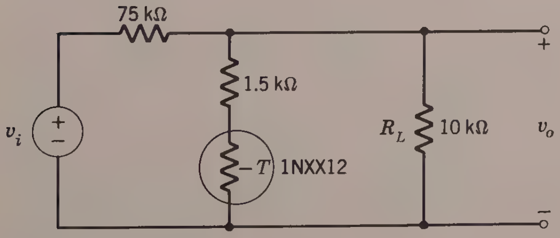
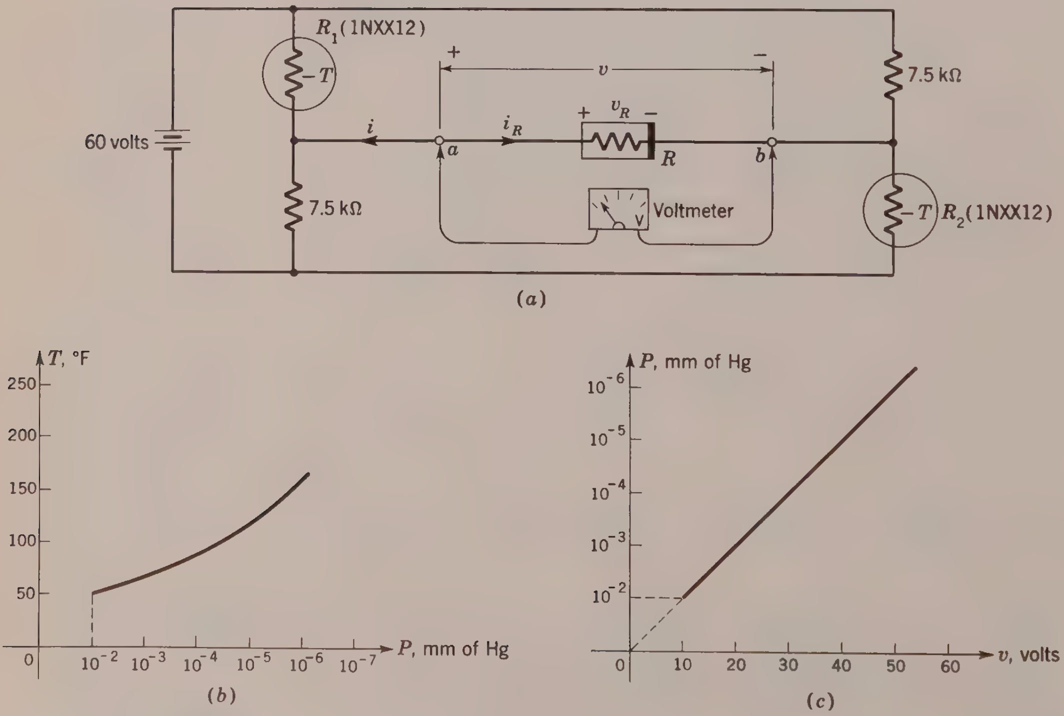


Fig. P9-15.

9-16 An accurate method for designing a vacuum gage in practice is to use the principle that the quantity of gas in a vacuum chamber changes the heat dissipation constant, and hence the ambient temperature, of a thermistor placed inside the chamber. A simple way to design such a vacuum gage is to connect two *identical* thermistors in the bridge circuit shown in Fig. P9-16*a*. Thermistor R_1 is sealed inside an evacuated glass bulb and placed inside the chamber. This is the reference element. Thermistor R_2 is exposed directly to the environment of the chamber. This is the sensing thermistor. The ambient temperature of the reference thermistor is given to be 50°F. The vacuum pressure (in millimeters of mercury) inside the chamber changes the ambient temperature of the sensing thermistor in accordance with the experimentally determined curve shown in Fig. P9-16*b*.

(a) Find the family of v-vs.-i DP plots across terminals a-b (replace resistor R by an open circuit) with the ambient temperature T as the parameter.

(b) Determine the v_R-i_R curve of the calibrating resistor R so that the voltmeter (with infinite internal resistance) reading will be a *linear* function of the vacuum pressure as shown by the calibrated scale in Fig. P9-16*c*.



(*a*)

(*b*)

(*c*)

Fig. P9-16.

10 SYNTHESIS OF AC-RESISTIVE NONLINEAR FUNCTIONAL NETWORKS: THE BLACK-BOX APPROACH

10-1 CHARACTERISTICS OF ELECTRONICALLY CONTROLLED SWITCHES

The functional networks presented in Chap. 9 are dc-resistive networks, and hence can be characterized either by a TC plot or by a DP plot for all time t. However, there are many important practical resistive functional networks which contain ac sources, and therefore cannot be characterized by a TC plot or a DP plot for all time t. The ac sources in these networks usually play the role of controlling signals, and these signals generally consist of stepwise waveforms. This means that during some interval of time, the network can be considered as a dc-resistive network which can be characterized by a TC plot or a DP plot for that interval. Such networks can be considered as controlled resistors whose TC plots or DP plots vary with time.

The most common controlled resistors are the *electronically controlled switches.* In its most general form, an electronically controlled switch is a two-port black box connected between a load and a source, as shown in Fig. 10-1*a*. The black box usually has one or more controlling signals $s_1(t), s_2(t), \ldots, s_n(t)$, so that for some specified combinations of the signal waveforms, the black box reduces to a *series switch* as shown in Fig. 10-1*b*, to a *shunt switch* as shown in Fig. 10-1*c*, or to a set of switches as shown in Fig. 10-1*d*. For convenience, we shall henceforth refer to these controlling signals as the *switching signals.* We shall consider first the series switch and then the shunt switch, leaving the more complicated configuration shown in Fig. 10-1*d* to the last. Our main motivation for studying electronically controlled switches is that they behave exactly like mechanical switches except that they can be turned on and off much more rapidly. Moreover, the switching operation can be achieved by remote control.

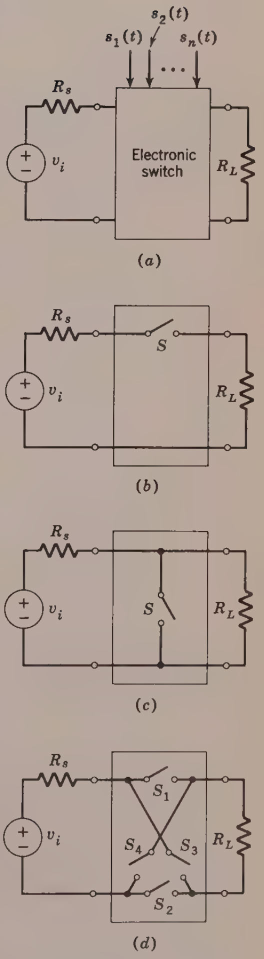


Fig. 10-1. A general configuration of an electronically controlled switch and its various specialized forms.

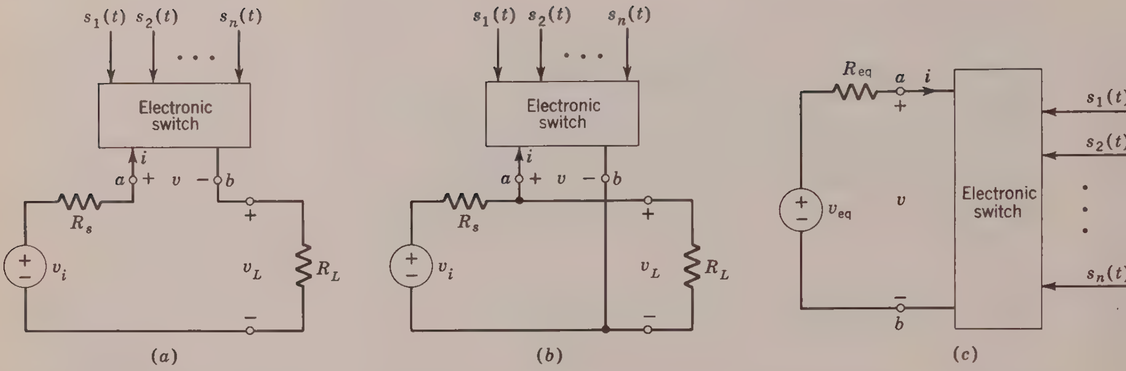


Fig. 10-2. The representation of a series switch and a shunt switch in equivalent forms.

It is more convenient to represent the electronically controlled switch as shown in Fig. 10-2*a* for the *series configuration,* or as shown in Fig. 10-2*b* for the *shunt configuration.* If we replace the network connected to terminals *a-b* of the electronic switch by its Thévenin equivalent circuit, then in both cases we obtain the equivalent network shown in Fig. 10-2*c*. For the series switch, we have

$$R_{\mathrm{eq}} = R_s + R_L \tag{10-1}$$

$$v_{\mathrm{eq}} = v_i \tag{10-2}$$

For the shunt switch, we have

$$R_{\mathrm{eq}} = \frac{R_s R_L}{R_s + R_L} \tag{10-3}$$

$$v_{\mathrm{eq}} = \frac{R_L}{R_s + R_L} v_i \tag{10-4}$$

Observe that as far as the electronic switch is concerned, the external network connected across its driving-point terminals is the same in both the series and the shunt configurations. It suffices, therefore, for us to characterize the electronic switch in terms of the equivalent network shown in Fig. 10-2*c*.

In order to function as a switch, it is necessary that for some specific combination of the switching signals, the electronic switch must "look like" an open circuit, and for another combination of the switching signals, the electronic switch must "look like" a short circuit. We intentionally use the term "look like" and not "become," because it is not necessary for the electronic switch to actually become an open circuit or a short circuit. In order to clarify these seemingly contradictory statements, let us refer to Fig. 10-2*c* again and observe that the function that must be performed by the electronic switch is this: given a range of values of v_{eq}, that is,

$$E_{\min} \leq v_{eq} \leq E_{\max} \tag{10-5}$$

we would like to have the current

$$i = 0 \tag{10-6}$$

for some specified combination of the switching signals, and to have the voltage

$$v = 0 \tag{10-7}$$

for another specified combination of the switching signals. Of course, if the electronic switch becomes either an open circuit or a short circuit, Eqs. (10-6) and (10-7) would automatically be satisfied. But this is not the only way these equations can be satisfied. In fact, if we draw the load line in the *v-i* plane corresponding to the two extreme values specified by Eq. (10-5), as shown in Fig. 10-3*a*, then at any time *t*, the load line corresponding to the input time function $v_{eq}(t)$ must necessarily be parallel to these load lines, and is bounded between them. Since at any time *t*, the voltage *v* and the current *i* are found to be the intersection between the DP plot of the electronic switch and the load line at time *t*, it is clear that in order to satisfy Eq. (10-6), it is both necessary and sufficient that the DP plot be given as shown in Fig. 10-3*b* by

$$i = \begin{cases} 0 & \text{for } E_{\min} \leq v \leq E_{\max} \\ \text{arbitrary elsewhere} \end{cases} \tag{10-8}$$

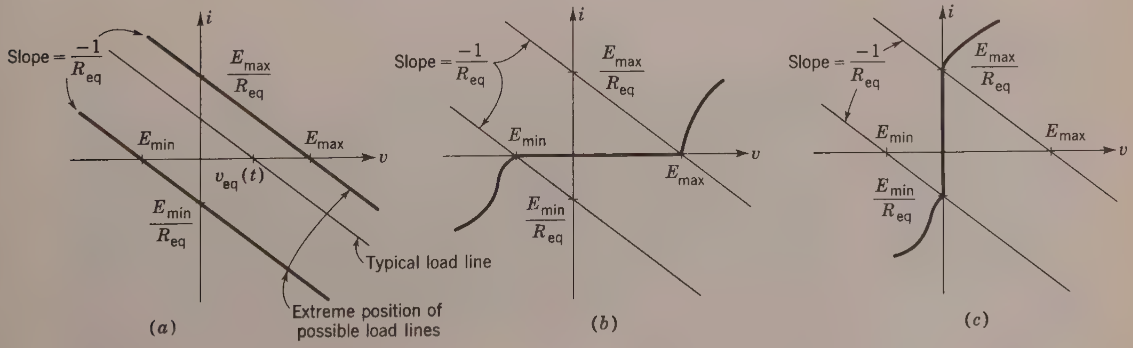


Fig. 10-3. The DP plots representing the two states of an electronically controlled switch.

Similarly, in order to satisfy Eq. (10-7), it is both necessary and sufficient that the DP plot be given as shown in Fig. 10-3*c* by

$$v = \begin{cases} 0 & \text{for } \dfrac{E_{\min}}{R_{\text{eq}}} \leq i \leq \dfrac{E_{\max}}{R_{\text{eq}}} \\ \text{arbitrary elsewhere} \end{cases} \tag{10-9}$$

The electronically controlled switch is, therefore, completely characterized by the two DP plots shown in Fig. 10-3*b* and *c* corresponding to the two specified combinations of switching signals. Observe that the DP plot shown in Fig. 10-3*b* "looks like" an open circuit only if the terminal voltage v lies within the interval $(E_{\min}, E_{\max})$. Similarly, the DP plot shown in Fig. 10-3*c* "looks like" a short circuit only if the terminal current i lies within the interval $(E_{\min}/R_{\text{eq}}, E_{\max}/R_{\text{eq}})$. For convenience, we shall say the switch is "off" if the switching signals result in $i = 0$, and we shall say the switch is "on" if the switching signals result in $v = 0$.

It must be emphasized that the two voltages $E_{\max}$ and $E_{\min}$ are associated with the equivalent voltage v_{eq} and not the actual input voltage v_i. However, given the range of the input voltages v_i, we can immediately determine the corresponding range of v_{eq} from Eq. (10-2) for the series switch or Eq. (10-4) for the shunt switch. Similarly, the value of R_{eq} can be determined from Eqs. (10-1) or (10-3). For convenience, we shall usually refer to v_{eq} as the *equivalent input signal.*

In many practical applications, the input voltage v_i assumes only one polarity for all time, that is, either $v_i(t) \geq 0$ or $v_i(t) \leq 0$. Correspondingly, we have either $E_{\min} = 0$ or $E_{\max} = 0$. The above type of electronically controlled switch will henceforth be referred to as a *unidirectional switch,* in contrast with a *bidirectional switch* where neither $E_{\min}$ nor $E_{\max}$ is zero. We shall now proceed to synthesize some commonly encountered electronically controlled switches.

10-2 SYNTHESIS OF SINGLE-CONTROLLED UNIDIRECTIONAL SWITCHES

Consider the synthesis of an electronic switch corresponding to the case where the equivalent input signal $v_{\text{eq}}(t) \geq 0$ for all time t, and where there is only one switching signal $s(t)$.[1] We would like the switch to be "on" when $s(t) = 0$ and "off" when $s(t) = E$. If the range of the equivalent input signal is

$$0 \leq v_{\text{eq}}(t) \leq E_{\max} \tag{10-10}$$

---

[1] It is important to differentiate between the equivalent input signal and switching signal. The equivalent input signal is the information-bearing signal, in contrast with the switching signal, whose only function is to turn on or off the electronic switch.

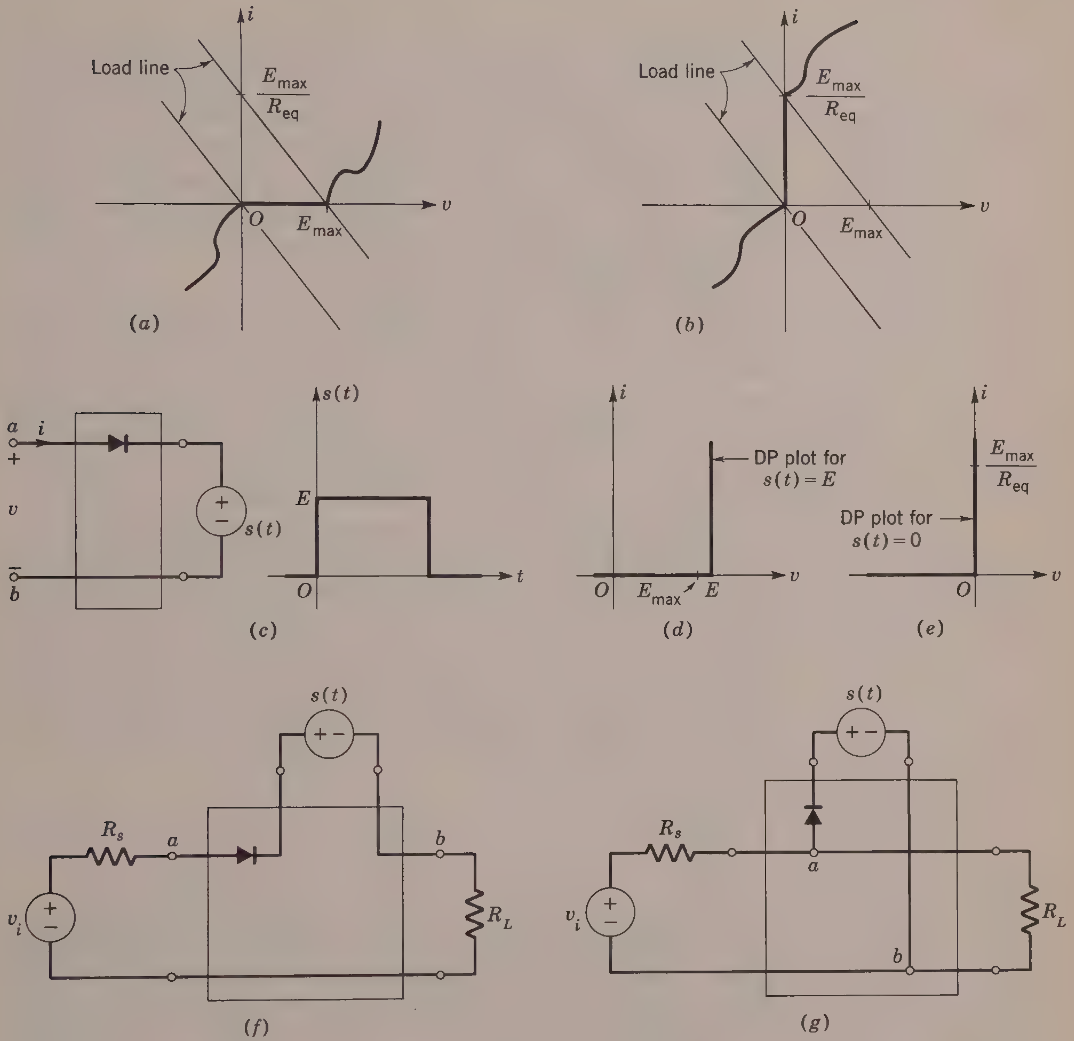


Fig. 10-4. The DP plots representing the two states of a single-controlled unidirectional switch and their realizations in the form of a series switch and a shunt switch.

then the DP plot of the switch must be as shown in Fig. 10-4*a* when $s(t) = E$ and in Fig. 10-4*b* when $s(t) = 0$. Observe that the *v-i* curve beyond the horizontal or vertical segment is completely arbitrary. This gives us freedom to choose from a number of components that might fit the above description. For example, the simple ideal diode circuit shown in Fig. 10-4*c* would qualify because when $s(t) = E$, the DP plot shown in Fig. 10-4*d* contains the prescribed horizontal segment, provided

$$E \geq E_{\max} \tag{10-11}$$

Similarly, when $s(t) = 0$, the DP plot shown in Fig. 10-4*e* contains the prescribed vertical segment. If we connect this circuit in place of the black box shown in Fig. 10-2*a* or *b*, we obtain the simple series switch shown in Fig. 10-4*f* or the shunt switch shown in Fig. 10-4*g*. Remember that these two circuits are applicable only when $v_i(t) \geq 0$. However, we can apply the complementary-network theorem (Sec. 7-8) to realize the corresponding circuit when $v_i(t) \leq 0$ by simply interchanging the two terminals of the ideal diode and the switching voltage source $s(t)$.

In practice, we must choose an element which closely resembles the ideal diode, and the junction diode is usually chosen for this purpose. We must expect, therefore, that a physical electronic switch can only approximate the ideal case. In other words, when the switch is "off," the current i in the switch will be quite small but not zero. Similarly, when the switch is "on," the voltage v across the switch will be quite small but not zero.

Although this switch is used in many applications, it has the disadvantage that in view of Eq. (10-11), a relatively large switching signal is required to turn the switch off; that is, the voltage must be at least as large as the maximum equivalent input voltage. In many remote-control systems, the switching signal must be transmitted over vast distances, and hence it is uneconomical to use a large switching signal. In this respect, it is convenient to define the *switching sensitivity* η of the switch as[1]

$$\eta = \frac{E_{\max}}{E} \tag{10-12}$$

where $E_{\max}$ is the maximum amplitude of the input signal and E is the minimum amplitude of the switching signal required for satisfactory operation. For example, the switching sensitivity of the switch presented above is equal to unity. Clearly, when the switching signal amplitude is our main concern, a high-switching-sensitivity switch would be desirable.

10-2-1 A HIGH-SWITCHING-SENSITIVITY TRANSISTOR SWITCH

The typical collector-to-emitter characteristic curves of an *n-p-n* transistor are shown in Fig. 10-5*a*. If we operate this transistor as a controlled resistor as shown in Fig. 10-5*b*, then the transistor *v-i* curve corresponding to the switching current $i_s(t) = I_B = 0$ as shown in Fig. 10-5*c* is a good approximation to the prescribed horizontal segment (representing an open circuit). Similarly, the

---

[1] Although the switching sensitivity is defined here as the ratio between two voltages, it could also be defined as the ratio between any two appropriate quantities, e.g., two currents.

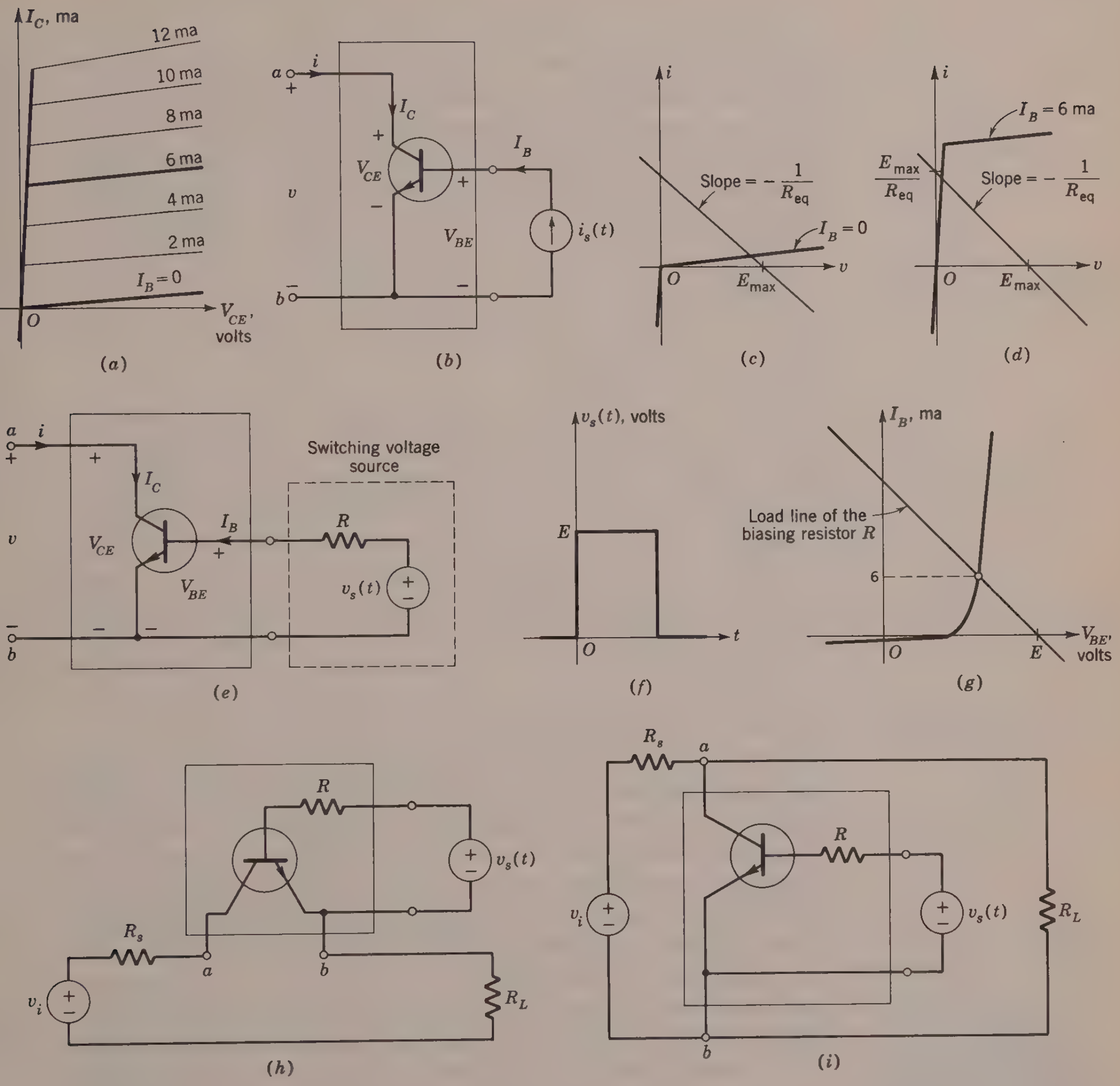


Fig. 10-5. The step-by-step procedure for synthesizing a series or shunt transistor switch.

transistor v-i curve corresponding to $i_s(t) = I_B = 6$ ma as shown in Fig. 10-5d is a good approximation to the prescribed vertical segment (representing a short circuit).[1] The circuit shown in Fig. 10-5b, therefore, qualifies as a unidirectional switch. Since it is inconvenient to use a current source as the switching signal, let us replace it by the more practical switching voltage source circuit shown in Fig. 10-5e. Corresponding to the voltage switching signal $v_s(t)$ in Fig. 10-5f, the value of the biasing resistor R is found by the load-line construction as shown in Fig. 10-5*g*. For most switching transistors, the value of E is a few volts. On the other hand, the value $E_{\max}$ may assume any prescribed value, say 30 volts, and hence, the switching sensitivity η for this circuit is much higher than for the diode switch. If we connect this circuit in place of the black box in Fig. 10-2*a* or *b*, we obtain the *series transistor switch* shown in Fig. 10-5*h* or the *shunt transistor switch* shown in Fig. 10-5*i*.

---

[1] Transistors which are designed specifically with an almost horizontal v-i curve when $I_B = 0$ and an almost vertical v-i curve portion when $I_B > 0$ are called *switching transistors.* These are readily available commercially.

10-2-2 A HIGH-SWITCHING-SENSITIVITY SCR LATCHING SWITCH

One disadvantage of the transistor switch presented above is that the switching voltage signal $v_s(t) = E$ must be applied at all times when the switch is to simulate a short circuit. This could cause considerable drain in power over long periods of time. It is, therefore, desirable to design a switch which will stay "closed" even after the switching signal is removed. This is called a *latching switch.* A switch which exhibits this property is given by the *silicon-controlled rectifier* (SCR) circuit shown in Fig. 10-6*a*.[1] The *v-i* curves of the SCR as a function of the controlling gate current I_G are shown in Fig. 10-6*b*. Observe that the required horizontal segment can be approximated by that portion of the *v-i* curve near the v axis when $I_G = 0$. Similarly, the required vertical segment can be approximated by that portion of the *v-i* curve near the i axis. The fundamental feature of the SCR *v-i* curves which makes it a latching device is that the near-vertical segment in Fig. 10-6*b* is a part of the *v-i* curve for all values of gate current. Hence, if we happen to be operating on this segment, then the operating point will remain on this segment even if the gate current is reduced to zero.

In order to analyze how the switching is effected, let us assume that initially we are operating on the horizontal segment corresponding to $I_G = 0$ as shown in Fig. 10-6*c*. If we further assume that the value of the equivalent voltage source is bounded by

$$E_{\min} < v_{\text{eq}} < E_{\max} \tag{10-13}$$

where $E_{\min}$ corresponds to the voltage intercept of the load line passing through point H of the SCR *v-i* curve, then at any time t the load line corresponding to $v_{\text{eq}}(t)$ will lie between the two extreme load lines shown in Fig. 10-6*c*. Accordingly, the operating

---

[1] The silicon-controlled rectifier is available commercially under many different trade names. Some of the more common aliases are the silicon-controlled switch (SCS), Trigistor, Thyristor, Transwitch, and Binistor. A slight distinction is sometimes given to these names in terms of the power-handling capability and the switching speed of the devices.

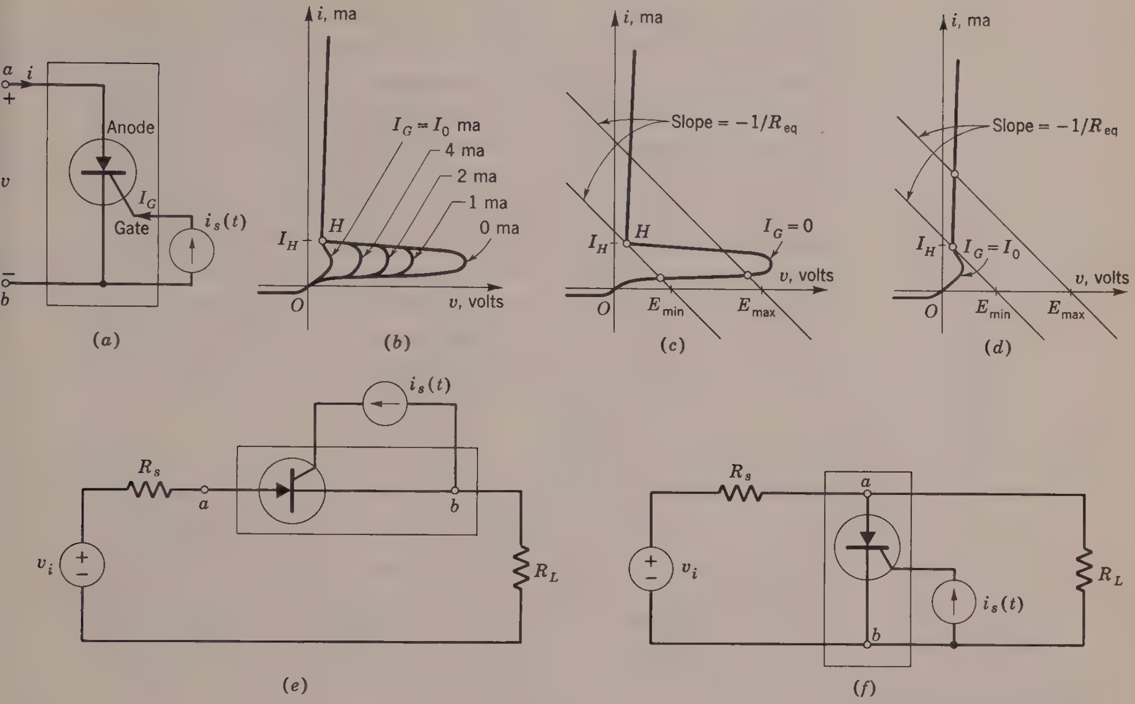


Fig. 10-6. Graphical analysis of an SCR latching switch.

point will always be on the lower horizontal segment.[1] Hence, the above situation corresponds to a switch in its "off" state. In order to turn on the switch, it is necessary to apply a large-enough switching current I_G so that the load lines intersect the resulting *v-i* curve only in its vertical segment, as shown in Fig. 10-6*d*. Once the switch is turned on and the controlling signal is removed, the *v-i* curve of the SCR switches back to the one shown in Fig. 10-6*c*. However, since the vertical segment of this *v-i* curve remains intact, the operating point will also remain intact, and the switch will remain "on."[2] The result is therefore a latching switch.

The only way we can turn off the switch is to force the load line shown in Fig. 10-6*c* to move below point *H* of the *v-i* curve, thus forcing the operating point to switch to the lower horizontal segment. In view of the significance of the point *H*, the corresponding current I_H is called the *holding current.* Hence, in order to turn off the switch, it is necessary to reduce the anode current *i* below the holding current, that is,

$$i < I_H \tag{10-14}$$

---

[1] Although strictly speaking there are three operating points in Fig. 10-6*c*, only the operating point along the lower horizontal segment is actually attained in practice. The reason for this will be obvious after Chaps. 13 and 14.

[2] Again, when $I_G = 0$, there will be three possible operating points, but in this case only the operating point along the vertical segment is actually attained in practice. This, too, will be obvious after Chaps. 13 and 14.

This can be done in a number of ways. The simplest way consists of disconnecting the anode from terminal a momentarily, thus forcing $i = 0$. Another way is to increase the value of R_{eq} momentarily or to decrease the value of v_{eq} below E_{min} so that the load line through $v = E_{min}$ will momentarily fall below point H.†

If we connect this switch in place of the black box shown in Fig. 10-2*a* or *b*, we would obtain the series latching switch shown in Fig. 10-6*e* or the shunt latching switch shown in Fig. 10-6*f*. In practice, the controlling gate current is usually obtained by a circuit similar to that shown in Fig. 10-5*e*. The required switching voltage is usually very small, and hence the switching sensitivity of this latching switch is extremely high. Because of this, the SCR is widely used in practice as a latching switch.

Exercise 1: Discuss the procedure for designing a transistor unilateral switch corresponding to the case $v_i(t) \leq 0$.

Exercise 2: Synthesize a unilateral switch using the FET as the switching element.

Exercise 3: In practice, the current source shown in Fig. 10-6*a* is generally replaced by an appropriate resistor-voltage source combination. Explain how this circuit can be designed. What additional information do you need?

Exercise 4: Synthesize a latching switch using the thyratron listed in Table 3-1 as the switching element.

Exercise 5: The *light-activated switch* (LAS) is a newly discovered light-controlled nonlinear resistor whose characteristic curves are similar to the SCR but with the gate current parameter replaced by light illumination. (*a*) Design a latching switch using this element. (*b*) Give some practical applications of this particular switch.

10-3 SYNTHESIS OF SINGLE-CONTROLLED BIDIRECTIONAL SWITCHES

Consider the more general case of an electronic switch where the equivalent input signal v_{eq} may assume both polarities; that is, $E_{max} > 0$ and $E_{min} < 0$. We shall consider here only one switching signal $s(t)$ so that the switch is on when $s(t) = 0$ and off when $s(t) = E$.‡ In this case, the electronic switch is characterized by the DP plot shown in Fig. 10-3*b* when $s(t) = E$, and in Fig. 10-3*c* when $s(t) = 0$. For convenience, let us, without loss of generality, assume that $E_{max} = -E_{min}$.§ In this case, the DP plots in Fig. 10-3*b* and *c* are symmetrical with respect to the origin over the region of interest. Hence, the symmetric DP plot theorem (Sec. 7-7-2) suggests that we connect two identical unidirectional switches (such as the diode switch in Fig. 10-4*c*) back to front in parallel as shown in Fig. 10-7*a*. The DP plots of this combination corre-

---

† Because of some physical phenomenon analogous to inertia in physical systems, it takes a finite time interval to turn on or to turn off an electronic switch. The turn-on and turn-off times due to inertia are usually furnished by the manufacturer of the device.

‡ Actually, the specific switching level of $s(t)$ is not important here. We could just as well require the switch to be on when $s(t) = E$ and off when $s(t) = 0$. In fact, any two distinct levels $s(t) = E_1$ and $s(t) = E_2$ are acceptable provided such a switching signal is available.

§ If this is not the case, we could simply set E_{max} and $-E_{min}$ equal to the larger of the two magnitudes. In this case, the required DP plots will become more conservative than is necessary.

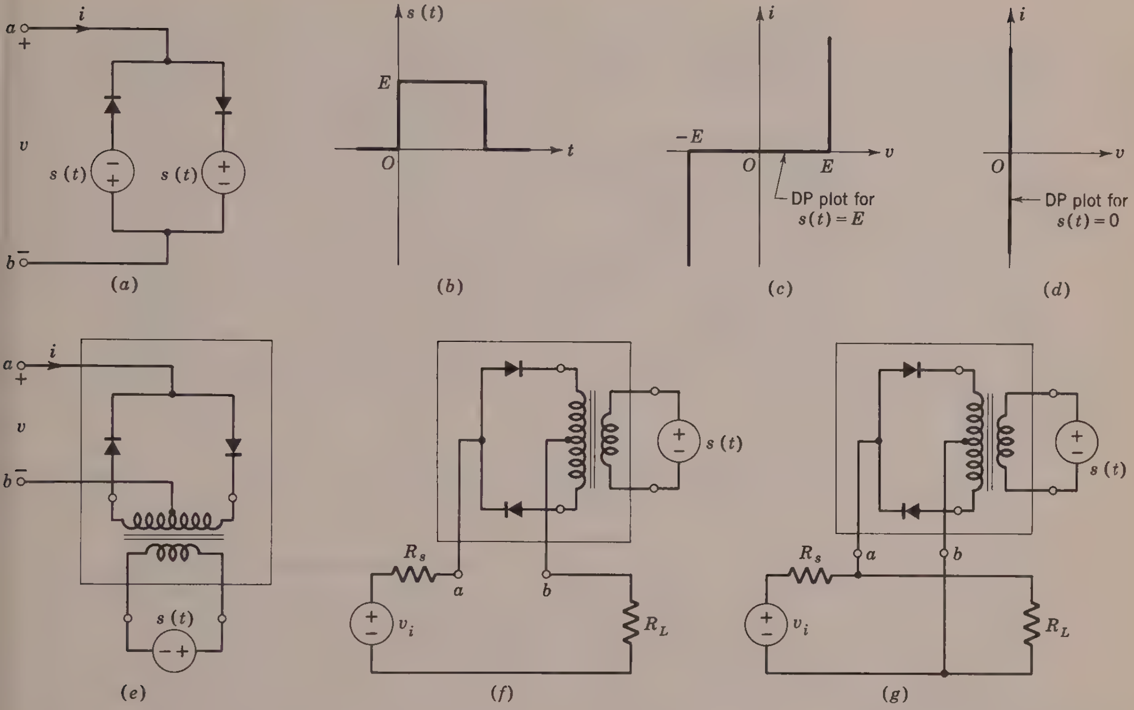


Fig. 10-7. Synthesis of a bidirectional electronic switch.

sponding to the switching signal in Fig. 10-7*b* are shown in Fig. 10-7*c* and *d*. Hence, if we choose the switching voltage level $E \geq E_{\max}$, then the DP plot in Fig. 10-7*c* would satisfy the requirements of Fig. 10-3*b*. The DP plot in Fig. 10-7*d* also satisfies the requirements of Fig. 10-3*c*, regardless of the equivalent input signal. Hence, the circuit shown in Fig. 10-7*a* is a bidirectional switch. One disadvantage of this circuit is that two identical switching sources are required. We may be tempted to apply the v-shift theorem here and replace them by a single source. However, we cannot do this because the two sources are connected back to front in series and therefore have the wrong polarity combination. However, these two sources may be replaced by an equivalent three-port transformer voltage-source combination (see Fig. 7-21*e*) as shown in Fig. 10-7*e*.[1] If we substitute this bidirectional switch in place of the black box in Fig. 10-2*a* or *b*, we obtain the bidirectional series switch shown in Fig. 10-7*f* and the bidirectional shunt switch shown in Fig. 10-7*g*, respectively.

Since physical transformers are bulky and heavy in comparison with solid-state devices, it is desirable to eliminate them if possible.[2] A closer examination of the DP plots in Fig. 10-7*c* and *d* might

[1] The transformers used throughout this book are assumed to be ideal for convenience. A physical transformer will deteriorate in performance as the operating frequency increases to a very high value or decreases to a very low value. In particular, an iron-core transformer will not function at all with direct current.

[2] In fact, based on the current state of the art of integrated-circuit technology, transformers are all but outlawed in virtually every circuit.

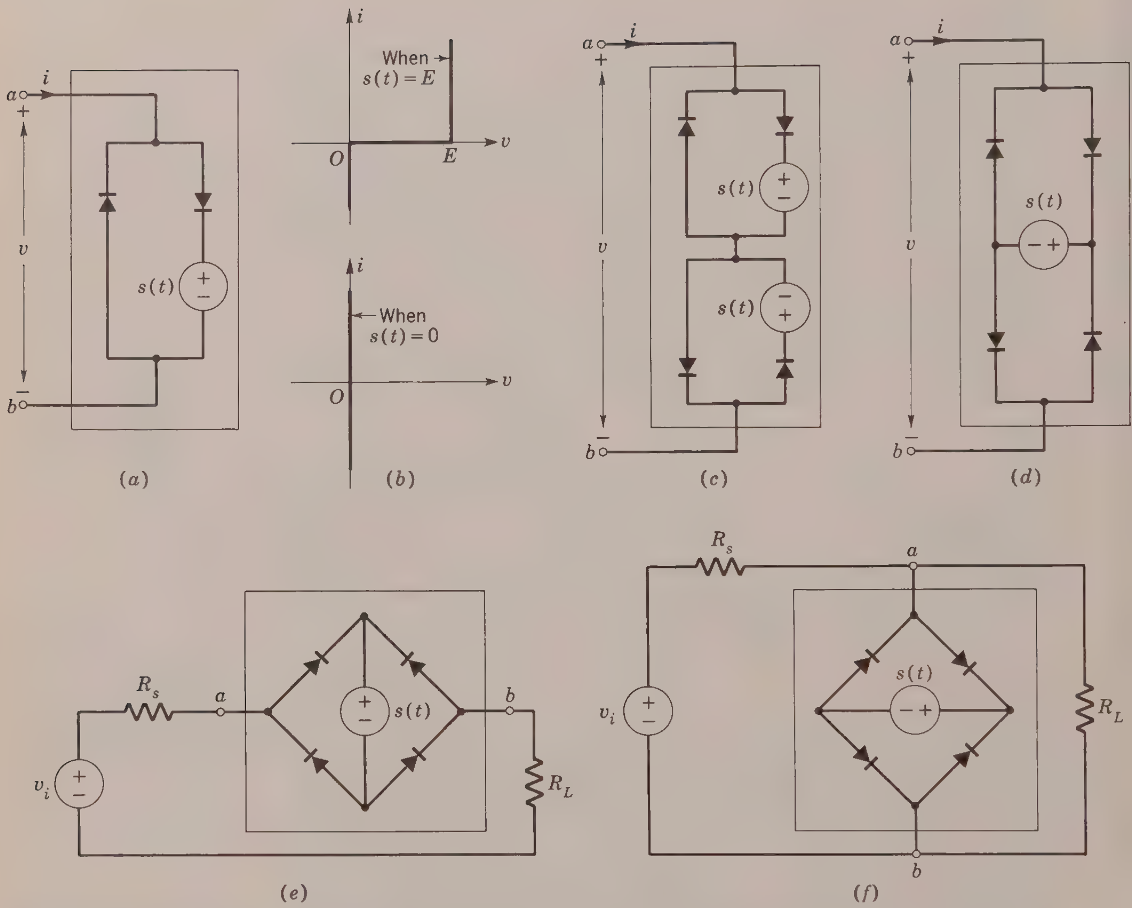


Fig. 10-8. An alternate realization of a transformerless bidirectional switch.

suggest an alternate circuit. For example, consider first the network shown in Fig. 10-8*a* whose DP plot is shown in Fig. 10-8*b*. If we connect two of these networks back to back in series as shown in Fig. 10-8*c*, we would obtain the same DP plot shown in Fig. 10-7*c* when $s(t) = E$, and in Fig. 10-7*d* when $s(t) = 0$. Therefore, the network shown in Fig. 10-8*c* is equivalent to the bidirectional switch in Fig. 10-7*a*. Notice that although two switching voltage sources are still required, they now have the proper polarity combination for a fruitful application of the v-shift theorem. The resulting equivalent bridge circuit shown in Fig. 10-8*d* constitutes a more practical realization. If we substitute this bridge network in place of the black box shown in Fig. 10-2*a* or *b*, we obtain the bidirectional series switch shown in Fig. 10-8*e*, or the bidirectional shunt switch shown in Fig. 10-8*f*. Comparing the circuits shown in Figs. 10-7 and 10-8, we observe that, in effect, we have traded the transformer for two diodes. This tradeoff is usually made in practice because transformers are generally more expensive and heavier than diodes.

Following the same procedure as before, we can easily synthesize many other bidirectional switches using transistors, SCRs, or other nonlinear resistors which are either available now or yet to be invented.[1] As a matter of fact, at least a few dozen electronic switches exist in various different configurations and using different components. From the black-box point of view, however, they are all characterized by the same DP plot shown in Fig. 10-3*b* and *c*. There is no point, therefore, for us to study each of these circuits as a separate entity.

Exercise 1: Discuss the deviation in performance of the transformerless bidirectional switch in Fig. 10-8 from the ideal switch when the ideal diodes are replaced by junction diodes.

Exercise 2: Synthesize a bidirectional electronic switch using transistors as the switching elements. You may use both *p-n-p* and *n-p-n* transistors. Avoid transformers if possible, but in any case use only one switching voltage source. Compare the switching sensitivity of this switch with the diode switch in Fig. 10-8.

Exercise 3: Repeat Exercise 2 but using FETs as the switching elements.

Exercise 4: Repeat Exercise 2 but using SCRs as the switching elements.

Exercise 5: A paradox arises in analyzing the bidirectional switching circuit shown in Fig. 10-8 when the switching signal $s(t)$ is allowed to take on negative values. Explain how this paradox can be resolved.

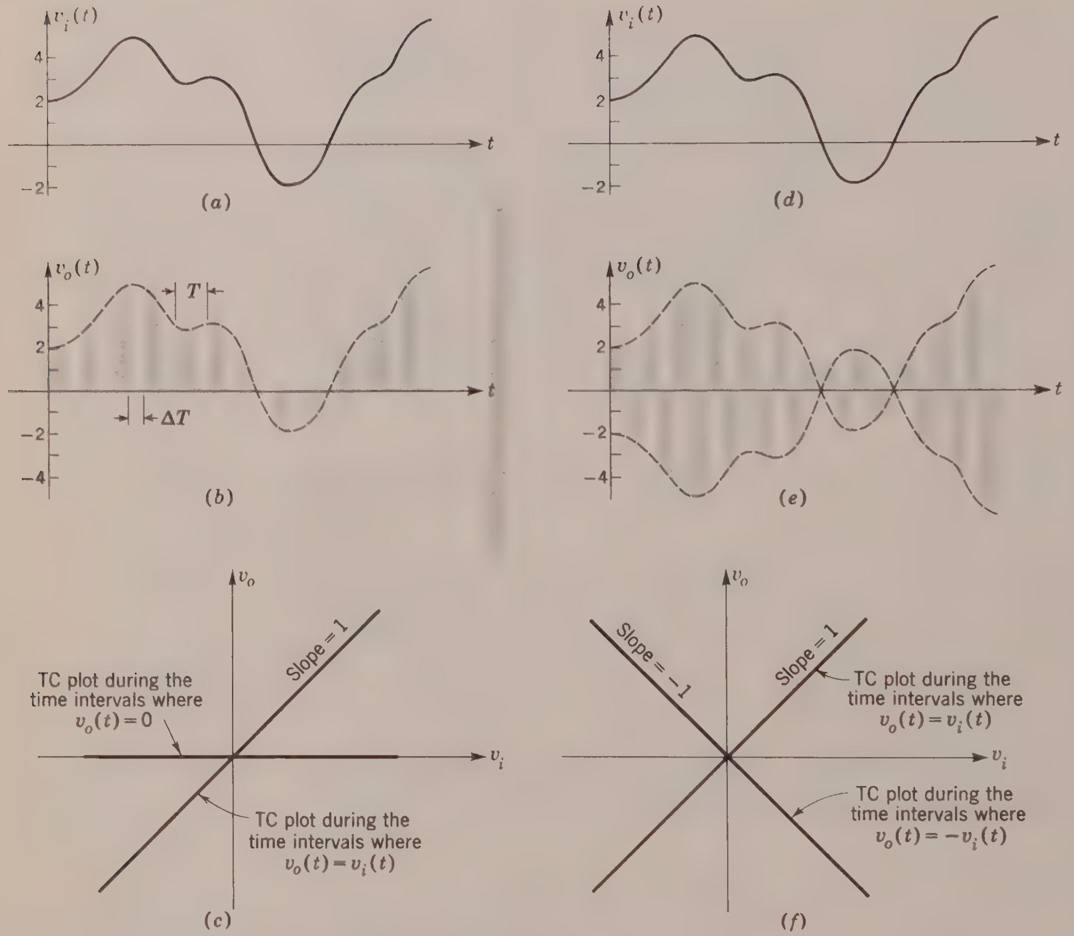
10-4 SYNTHESIS OF CHOPPERS AND AMPLITUDE MODULATORS

There are two very common waveform operations which are essential in many communication systems. One of these operations consists of "chopping" up an input signal, such as the time function $v_i(t)$ shown in Fig. 10-9*a* and transforming it into the output time function shown in Fig. 10-9*b*. For obvious reasons, a black box that accomplishes this task is called a *chopper*. Because the function of a chopper is to take samples of the input signal at regular intervals of time, the black box is sometimes called a "sampling circuit." An inspection of Fig. 10-9*a* and *b* immediately suggests that a chopper can be characterized by the two linear TC plots shown in Fig. 10-9*c* during alternate intervals of time.

The second common operation is similar to the chopper except that every alternate "slice" of the input signal is inverted about the time axis as shown in Fig. 10-9*d* and *e*. The correspond-

---

[1] It should now be clear that any controlled resistor which possesses a near-horizontal v-i curve for some controlling voltage and a near-vertical v-i curve for another controlling voltage is a potential component for synthesizing electronic switches.



$v_i(t)$
4
2
-2
t
(a)
$v_o(t)$
T
ΔT
4
2
-2
-4
t
(b)
v_o
v_i
Slope = 1
TC plot during the time intervals where $v_o(t) = 0$
TC plot during the time intervals where $v_o(t) = v_i(t)$
(c)
$v_i(t)$
4
2
-2
t
(d)
$v_o(t)$
4
2
-2
-4
t
(e)
v_o
v_i
Slope = −1
Slope = 1
TC plot during the time intervals where $v_o(t) = v_i(t)$
TC plot during the time intervals where $v_o(t) = -v_i(t)$
(f)

Fig. 10-9. The chopper and the amplitude modulator can be characterized by two TC plots which are switched back and forth by a switching signal.

ing TC plots during alternate intervals of time are shown in Fig. 10-9*f*. A circuit which meets this specification is called an *amplitude modulator*.

10-4-1 PRACTICAL CHOPPERS

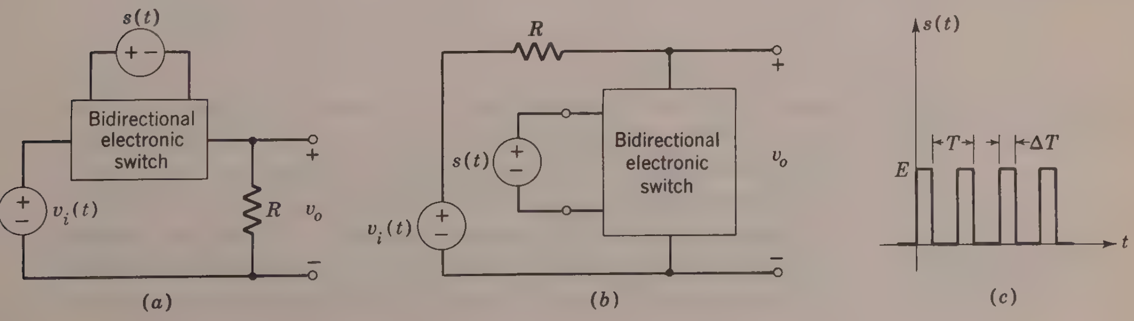
The TC plots shown in Fig. 10-9*c* can be realized by a bidirectional electronic switch connected either in series, as shown in Fig. 10-10*a*, or in shunt, as shown in Fig. 10-10*b*, with the input signal $v_i(t)$. The switching signal $s(t)$ as shown in Fig. 10-10*c* is chosen according to the prescribed width ΔT of each signal slice and the prescribed interval T between signal slices. In theory, any bidirectional electronic switch can be used for realizing the two chopper circuit configurations shown in Fig. 10-10*a* and *b*. In practice, we generally connect a resistor R_s in series with the input voltage source $v_i(t)$ of Fig. 10-10*a* and a resistor R_L across the output terminals of Fig. 10-10*b*. In either case, this results in a voltage divider, and the output voltage is, therefore, smaller than the input voltage by the voltage-divider ratio. For example, the networks shown in Figs. 10-7 and 10-8 are some typical realizations. In practice, the chopper circuit shown in Fig. 10-8*e* is usually called a *series modulator,* while the chopper circuit shown in Fig. 10-8*f* is usually called a *shunt modulator,* or a *Cowan modulator* after its inventor.

10-4-2 PRACTICAL AMPLITUDE MODULATORS

Conceptually, the simplest way to synthesize the TC plot shown in Fig. 10-9*f* is to use four bidirectional electronic switches, as represented symbolically in the circuit of Fig. 10-11*a* by S_1, S_2, S_3, and S_4. Each switch is to be controlled by the switching signal $s(t)$ shown in Fig. 10-11*b*, such that when $s(t) = E$, switches S_1 and S_2 are on while switches S_3 and S_4 are off, and when $s(t) = -E$, S_1 and S_2 are off while S_3 and S_4 are on. It is possible to eliminate two of these electronic switches by trading them for two transformers. This can be accomplished by either the network shown in Fig. 10-11*c* or the network shown in Fig. 10-11*d*.

Observe that the two electronic switches in each of the two alternate realizations shown in Fig. 10-11*c* and *d* require two separate switching signals. Our previous experience with the synthesis of electronic switches, however, suggests the possibility of reducing these to only one switching signal by means of the network shown in Fig. 10-11*e*. In this case, observe that when $s(t) = E$, diodes D_1 and D_2 become short circuits while diodes D_3 and D_4 become open circuits. Similarly, when $s(t) = -E$, diodes D_1 and D_2 become open circuits while diodes D_3 and D_4 become short cir-

Fig. 10-10. Two basic configurations of practical choppers.


s(t)
+ −
Bidirectional electronic switch
+
v_i(t)
R
v_o
−
(a)
R
+
s(t)
v_i(t)
Bidirectional electronic switch
v_o
−
(b)
s(t)
T
ΔT
E
t
(c)

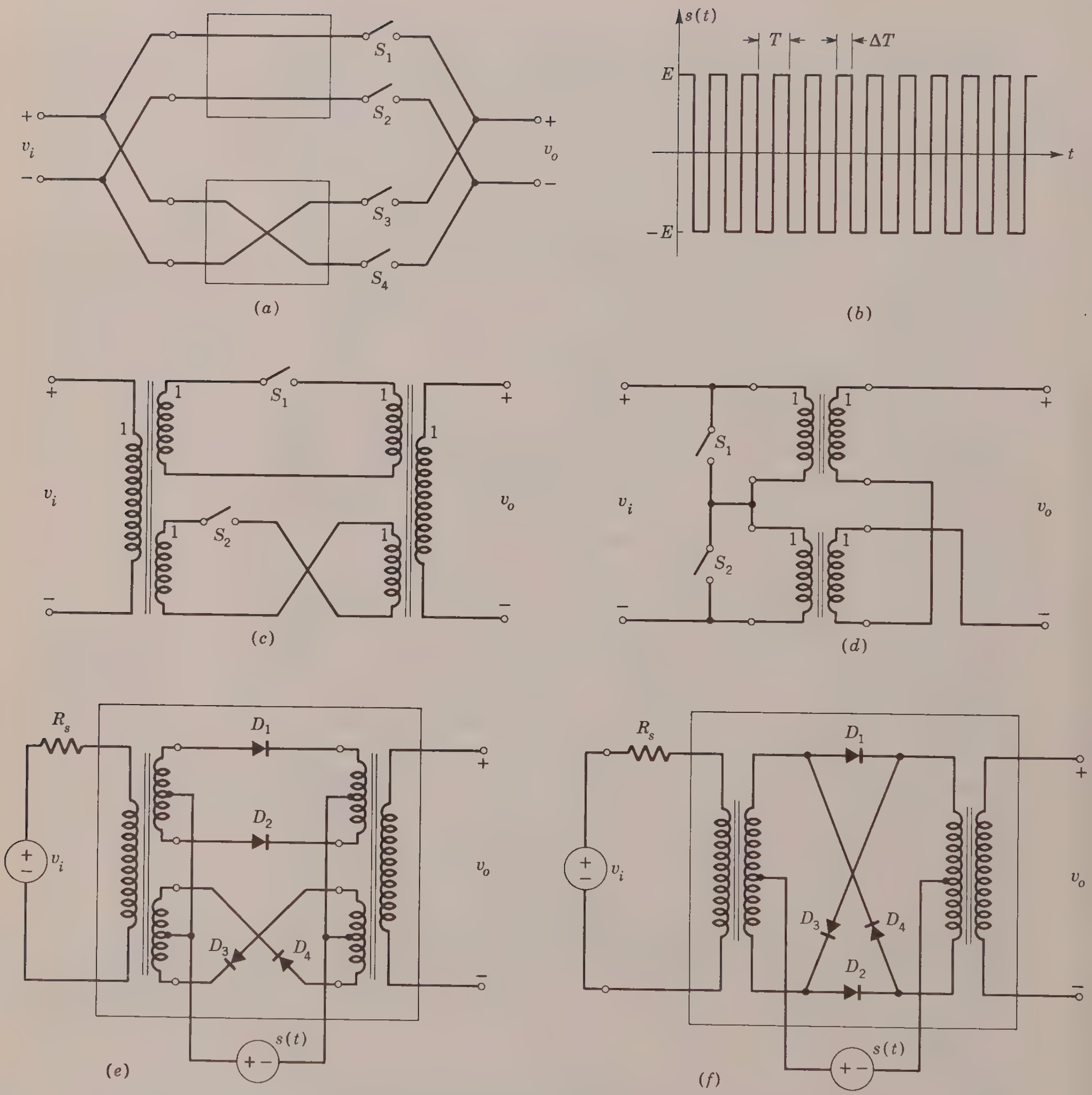


Fig. 10-11. The step-by-step procedure leading to the synthesis of a ring modulator.

cuits. Moreover, the two cases are independent of one another in the sense that only two of the diodes have an effect in the network at any one time. This observation suggests that if we combine the two sets of diodes as shown in Fig. 10-11*f*, we would still obtain the same result. The two networks shown in Fig. 10-11*e* and *f* are, therefore, equivalent. Observe that if we trace along the forward direction of each diode in Fig. 10-11*f*, we would eventually return to the original point. For example, starting with diode D_1, we proceed to D_3, then to D_2, then to D_4, and finally back to D_1. This gives the appearance of a ring, and hence the above network is commonly called a *ring modulator.*

Both the chopper and the amplitude modulator are widely used in telemetering equipment and other long-distance communication systems. Some of these applications are given in the form of problems at the end of this chapter.

Exercise 1: Discuss the disparity between the TC plots of the practical chopper circuits in Fig. 10-8 and the ideal TC plots shown in Fig. 10-9*c*.

Exercise 2: Show that the transformers in the ring modulator can be replaced by four linear resistors. What are the disadvantages of doing this in terms of the disparity between the resulting TC plots and that shown in Fig. 10-9*f*?

Exercise 3: Find another circuit for realizing an amplitude modulator.

10-5 SYNTHESIS OF MULTICONTROLLED ELECTRONIC SWITCHES

So far we have considered only electronic switches which are turned on and off by a single switching signal $s(t)$. Let us now consider the more general case where a predesignated combination of several switching signals $s_1(t), s_2(t), \ldots, s_n(t)$ are needed to turn on and off the switch. We shall refer to this class of circuits as *multicontrolled switches.* While a single-controlled switch is analogous to a padlock, a multicontrolled switch is analogous to a combination lock. Just as the latter is much more versatile, the multicontrolled switch has a much wider field of application. Referring again to the general configuration shown in Fig. 10-2*c*, it is clear that the DP plot of a multicontrolled bidirectional switch is still given by Fig. 10-3, and that for a multicontrolled unidirectional switch is given by Fig. 10-4.

We shall consider only the synthesis of a multicontrolled unidirectional switch. The bidirectional case can be realized in a manner similar to that presented in Sec. 10-3. Consider what happens if we connect the two unidirectional switches in parallel as shown in Fig. 10-12*a*. Let us assume for simplicity that the switching signals $s_1(t)$ and $s_2(t)$ are stepwise waveforms as shown in Fig. 10-12*b* and *c*, respectively. The DP plots across each switch at any time $t = t_0$ are shown in Fig. 10-12*d* and *e*. The resulting DP plot is shown in Fig. 10-12*f*.[1] Observe that the breakpoint voltage $v = E$ at $t = t_0$ is always equal to the smaller of the two breakpoint voltages $s_1(t_0)$ and $s_2(t_0)$. In other words, the breakpoint voltage of the resulting DP plot at any time t depends on

---

[1] Review Sec. 6-3-1 on how to combine DP plots with vertical segments.

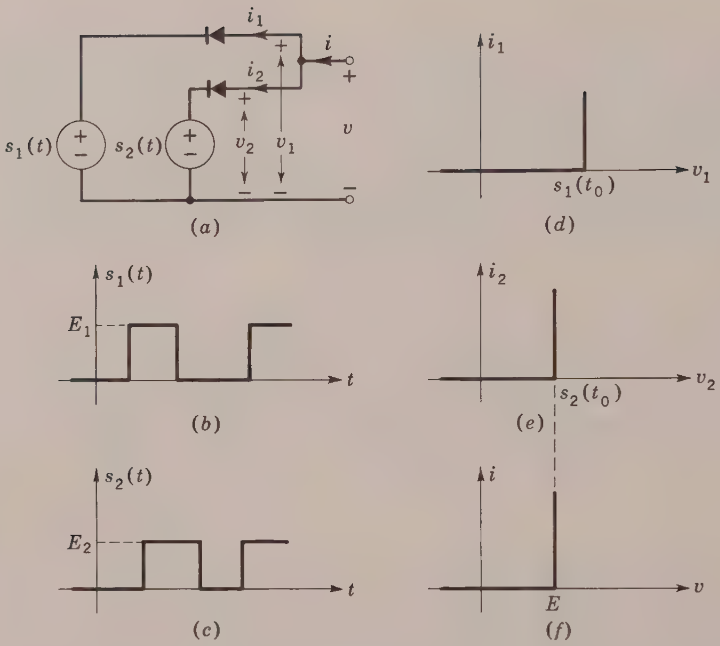


Fig. 10-12. The DP plot of two unidirectional switches in parallel has a breakpoint voltage equal to the smaller of the breakpoint voltages of the component DP plots.

the particular amplitude combination of $s_1(t)$ and $s_2(t)$ at the same time. More generally, if n sections of this unidirectional switch are connected in parallel with switching signals $s_1(t)$, $s_2(t)$, . . . , $s_n(t)$, then the DP plot at any time t is still given by Fig. 10-12*f*, where the breakpoint voltage E is equal to the smallest amplitude of these signals at the same time. Since the smallest amplitude of each switching signal is zero, the DP plot will look like a short circuit (switch is on) whenever one switching signal is zero at a given time. This is equivalent to saying that the resulting switch will open at a given time t if, and only if, all switching signals are simultaneously present (not zero) at the same time t. It should now be clear that this property allows us to choose all kinds of combinations to make the switch open or close at a given time. For example, this property can be used in the design of an electronic combination lock. Let us now consider two common applications.

10-5-1 SYNTHESIS OF A COINCIDENCE GATE

If we substitute an n-section multicontrolled switch in place of the black box shown in Fig. 10-2*c*, we would obtain the equivalent circuit shown in Fig. 10-13*a*.[1] Notice that if we let E_1, E_2, . . . , E_n

---

[1] To avoid unnecessary complexity in drawing, only three sections are shown.

be the amplitude of the respective switching signals, and if E is the smallest among these, then it is necessary that the equivalent input signal satisfies

$$0 \le v_{\text{eq}}(t) \le E \tag{10-15}$$

for all time t. Under this condition, the output voltage $v_o(t) = v_{\text{eq}}(t)$ if, and only if, none of the switching signals is zero at the same time. On the other hand, if at least one switching signal is zero, then $v_o(t) = 0$. The sample waveforms shown in Fig. 10-13*b* to *f* would illustrate this property more clearly. Since the input signal can arrive at the output terminals only if all switching signals coincide (not zero) at a given time, this circuit is generally known as a *coincidence gate.* We shall now show how coincidence gates can be combined to synthesize what is known as a transmission gate.

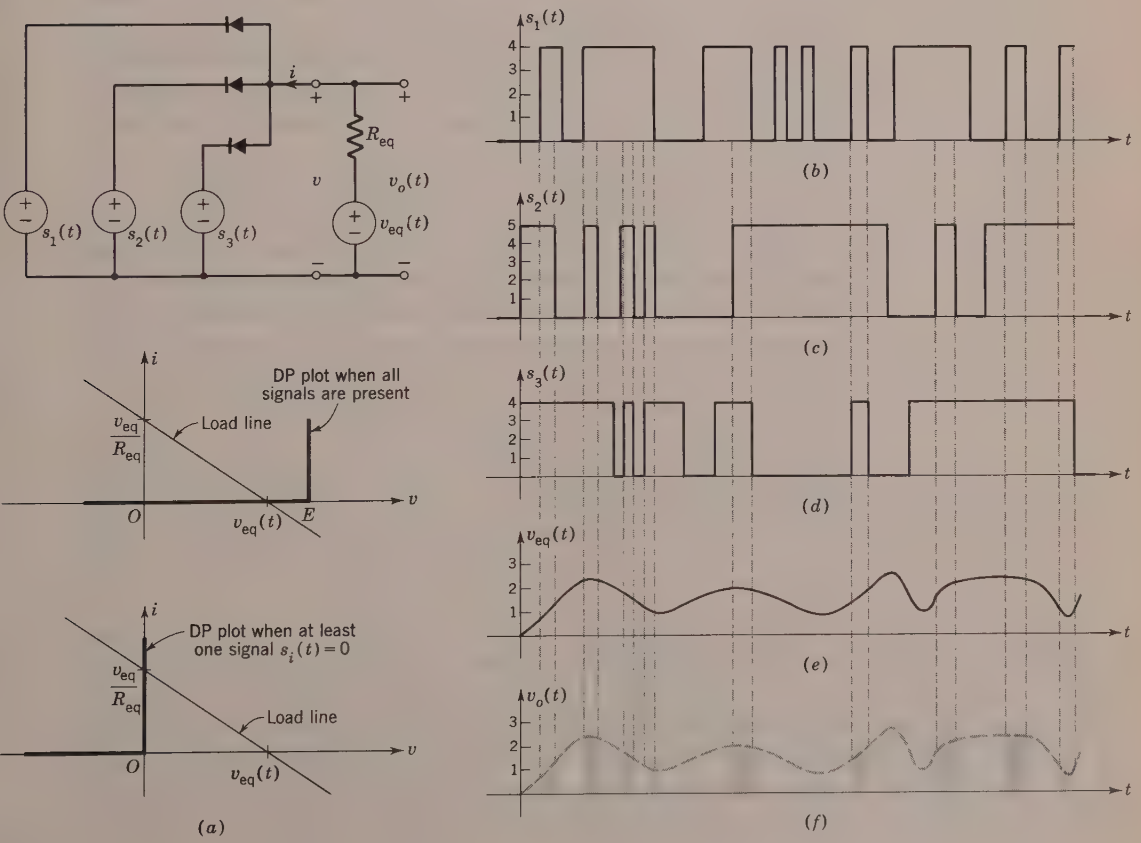


Fig. 10-13. The synthesis of a coincidence gate and a demonstration of its operating properties.

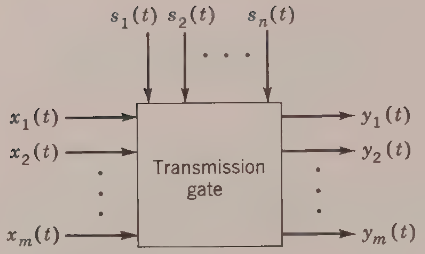


Fig. 10-14. Symbolic representation of a multi-input, multi-output transmission gate.

10-5-2 SYNTHESIS OF MATRIX TRANSMISSION GATES

Consider the multi-input multi-output black box shown in Fig. 10-14, where the input and output signals are represented by $x_1(t), x_2(t), \ldots, x_m(t)$ and $y_1(t), y_2(t), \ldots, y_m(t)$, respectively. The switching signals are given, as usual, by $s_1(t), s_2(t), \ldots, s_n(t)$. This black box is called a *transmission gate* if for some specified combination of switchings, a number of preselected input signals appear at the output terminals. For example, Fig. 10-15 shows a circuit made up of coincidence gates.[1] This is easily seen by applying the v-shift theorem to shift $s_2(t)$ into two identical but separate sources. One of the two coincidence gates is made up of diodes D_1 and D_2 and is controlled by $s_1(t)$ and $s_2(t)$. The other gate consists of diodes D_3 and D_4 and is controlled by $s_2(t)$ and $s_3(t)$.

Observe that aside from the fact that the two coincidence gates share one common switching signal $s_2(t)$, they are completely independent of each other. The three switching signals can be used to select any desired combination of output signals. For example, if we want $y_1(t) = x_1(t)$ and $y_2(t) = 0$, we simply apply the signals $s_1(t) = s_2(t) = E$ and $s_3(t) = 0$. If we want $y_1(t) = 0$ and $y_2(t) = x_2(t)$, we apply the signals $s_1(t) = 0$ and $s_2(t) = s_3(t) = E$. If we want both $y_1(t) = x_1(t)$ and $y_2(t) = x_2(t)$, we apply the signals $s_1(t) = s_2(t) = s_3(t) = E$. Finally, if we want $y_1(t) = 0$ and $y_2(t) = 0$, then we apply any combination of the switching signals other than those considered above. Hence, with three switching signals we can select four possible combinations of the output signals.

Clearly, by an extension of this idea, it is possible to select any number of input signals to appear at the output terminals by applying a suitable combination of switching signals to an appropriate interconnection of coincidence gates. When more than two coincidence gates are used, it is convenient to draw the network in a more systematic form. One common configuration consists of an array or matrix of diodes such as the network shown in

---

[1] To simplify the schematic diagram, it is common practice to omit drawing all wires connecting voltage sources or current sources to a common terminal, usually called the ground.

Fig. 10-16, where there are 10 input signals, 10 output signals, and eight switching signals. Observe that the diodes in each row of this diode matrix constitute a coincidence gate. Hence, there are a total of 10 coincidence gates in this diode matrix. An examination of the transmission gate shows that by a suitable choice of four out of the eight switching signals, we can select any combination of input signals. For example, to obtain $y_9(t) = x_9(t)$, we have to apply only the switching signals $s_2(t) = s_4(t) = s_6(t) = s_7(t) = E$. On the other hand, if we want to have $y_2(t) = x_2(t)$ and $y_{10}(t) = x_{10}(t)$ simultaneously, then we must apply $s_1(t) = s_4(t) = s_6(t) = s_7(t) = s_8(t) = E$.

Exercise 1: Find what happens if the switching signals in Fig. 10-12 are allowed to become negative. What is its effect on the coincidence gate?

Exercise 2: Synthesize a multi-input bidirectional switch.

Exercise 3: Show how a coincidence gate may be synthesized to function as a chopper.

Exercise 4: Synthesize a coincidence gate capable of transmitting an input signal of both polarities.

Exercise 5: Design an electronic combination lock to be used in a bank vault. You may use one or more matrix transmission gates.

10-6 SYNTHESIS OF LOGIC-CIRCUIT BUILDING BLOCKS

Logic can be defined as the science of correct thinking. It dispenses with human intuition and concerns itself with the truth and falsity of statements as judged strictly from the light of a given set of axioms which define the rules of the game. For example, if we let

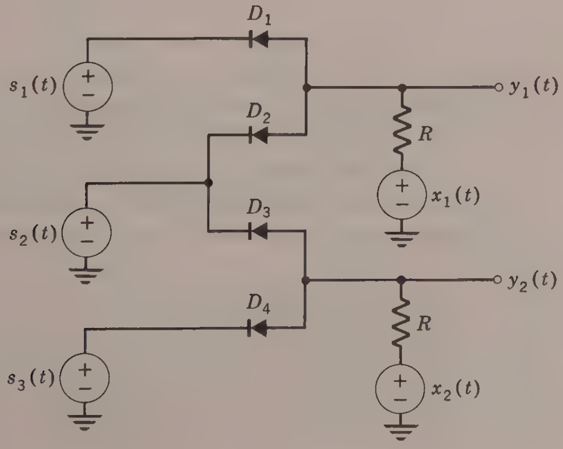


Fig. 10-15. An example of a transmission gate made up of two coincidence gates.

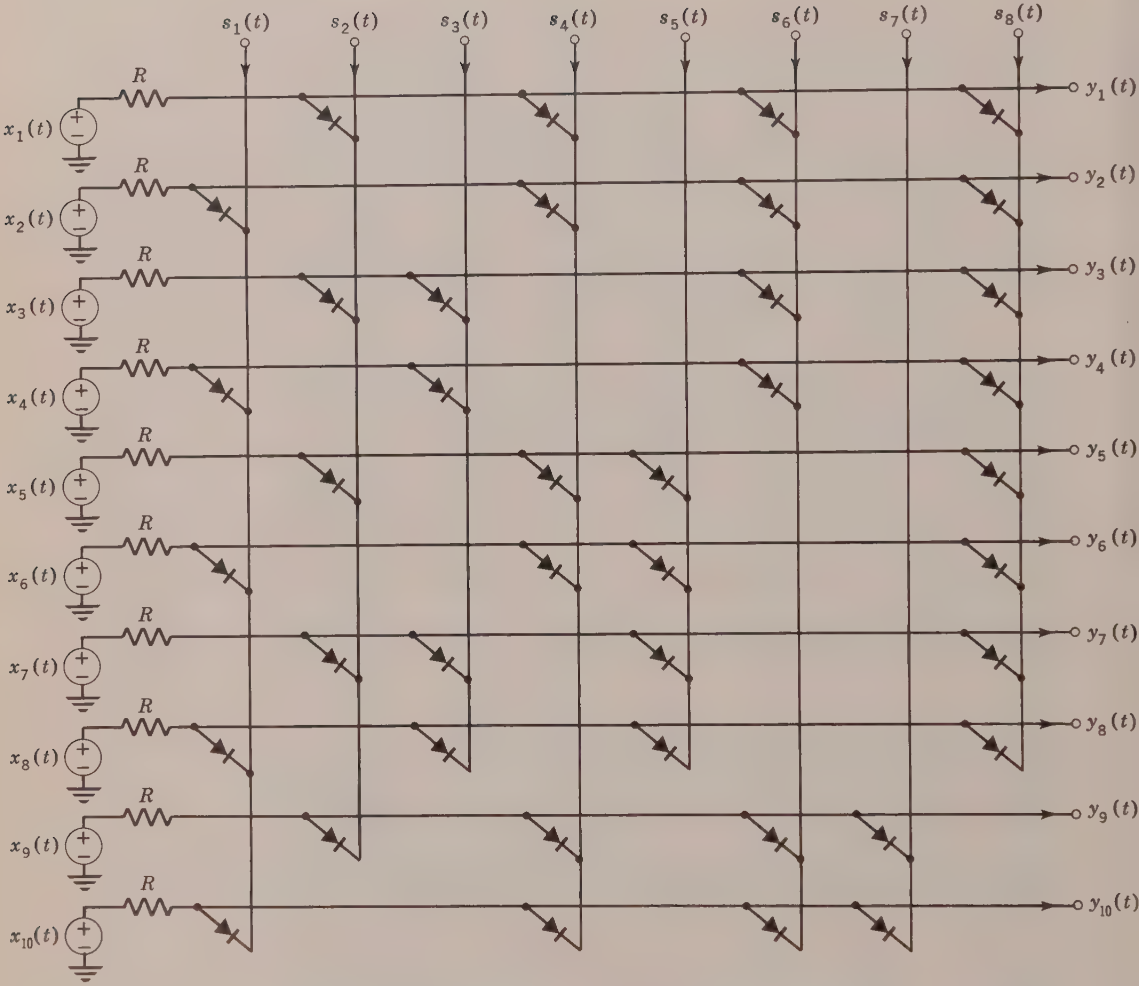


Fig. 10-16. An example of a matrix transmission gate containing 10 coincidence gates and eight switching signal sources.

x_1 and x_2 represent any two statements or propositions, each of which may either be true or false, and if we combine these statements to form a new statement y, then we may define a "rule" to the effect that statement y is true if, and only if, both statements x_1 and x_2 are true. Instead of stating this in so many words, we can express this rule by the following table:

x_1	x_2	y
True	True	True
True	False	False
False	True	False
False	False	False

Notice that this table is an equivalent and somewhat more precise way for stating the above rule. It is called a *truth table.* Instead of spelling out the words *true* and *false,* much time is saved by denoting them by symbols. It is customary to denote the word *true* by a 1 and the word *false* by a 0. It is important to realize that the 1 and the 0 are strictly symbols. They have no numerical significance whatsoever. Under this new notation, the above truth table becomes

x_1	x_2	y
1	1	1
1	0	0
0	1	0
0	0	0

This particular rule is called a *logic AND function*. Since it is one of the most commonly encountered functions, it is tabulated in Table 10-1 with the four other common logic functions, namely, the *logic OR function,* the *logic NAND function,* the *logic NOR function,* and the *logic NOT function.*[1] Using these five logic functions as basic building blocks, we can synthesize an infinite variety of more complex logic functions. Such logic functions are indispensable in the design of automatic machines with a certain amount of built-in artificial intelligence. For example, a logic function can be found to perform all possible alternatives in the routing of elevators. They can be used in place of the mechanical system of cams in an automatic washing machine. In fact, they are the basic building blocks of all digital systems, including digital computers. The study of how to find the logic functions for performing a given task is usually given in a course called *switching circuit theory*. Our objective in this section will be to study how the five basic logic functions can be implemented by nonlinear networks. Once we have these networks, then we can interconnect them to realize any logic function we need, provided we are careful to take the loading effects into consideration.[2]

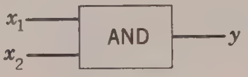
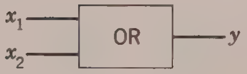
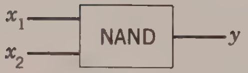
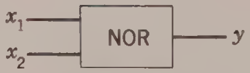
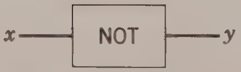
Since each input variable can assume only state 1 or state 0, we can easily represent these two states by two possible voltage or current levels. If we choose two convenient voltage levels E_1 and E_0 and associate the level E_1 with state 1 and level E_0 with state 0, then all we need to find are networks that will give an output voltage level corresponding to a combination of input voltage levels in accordance with the desired logic function. For reasons that will soon be obvious, the networks which realize the five basic logic functions in Table 10-1 are called AND gates, OR

---

[1] For simplicity, these functions are given only for one or two input variables; it should be recognized, however, that they (except for the NOT function) can be defined for any number of input variables.

[2] These five types of basic building blocks are now available commercially as black boxes with only the pertinent terminals brought out for external connections.

TABLE 10-1 *Five common logic functions and their defining truth tables.*

Logic function	Symbol	Truth table		
		x_1	x_2	y
AND		0	0	0
		0	1	0
		1	0	0
		1	1	1
		x_1	x_2	y
OR		0	0	0
		0	1	1
		1	0	1
		1	1	1
		x_1	x_2	y
NAND (NOT-AND)		0	0	1
		0	1	1
		1	0	1
		1	1	0
		x_1	x_2	y
NOR (NOT-OR)		0	0	1
		0	1	0
		1	0	0
		1	1	0
		x	y	
NOT		0	1	
		1	0	

gates, NAND gates, NOR gates, and NOT gates. Since we are concerned only with stepwise signals which can assume one of two levels, E_1 or E_0, at any given time, we can apply the results in the preceding sections to realize the above logic gates. Again, for simplicity, we shall restrict our discussions to one or two input variables only. The extension to n input variables is quite obvious.

10-6-1 SYNTHESIS OF AND GATES

From the truth table for the AND gate in Table 10-1, we find a network which produces an output voltage $v_o(t) = E_1$ if, and only if, both input signals $v_1(t) = E_1$ and $v_2(t) = E_1$. Otherwise $v_o(t) = E_0$.

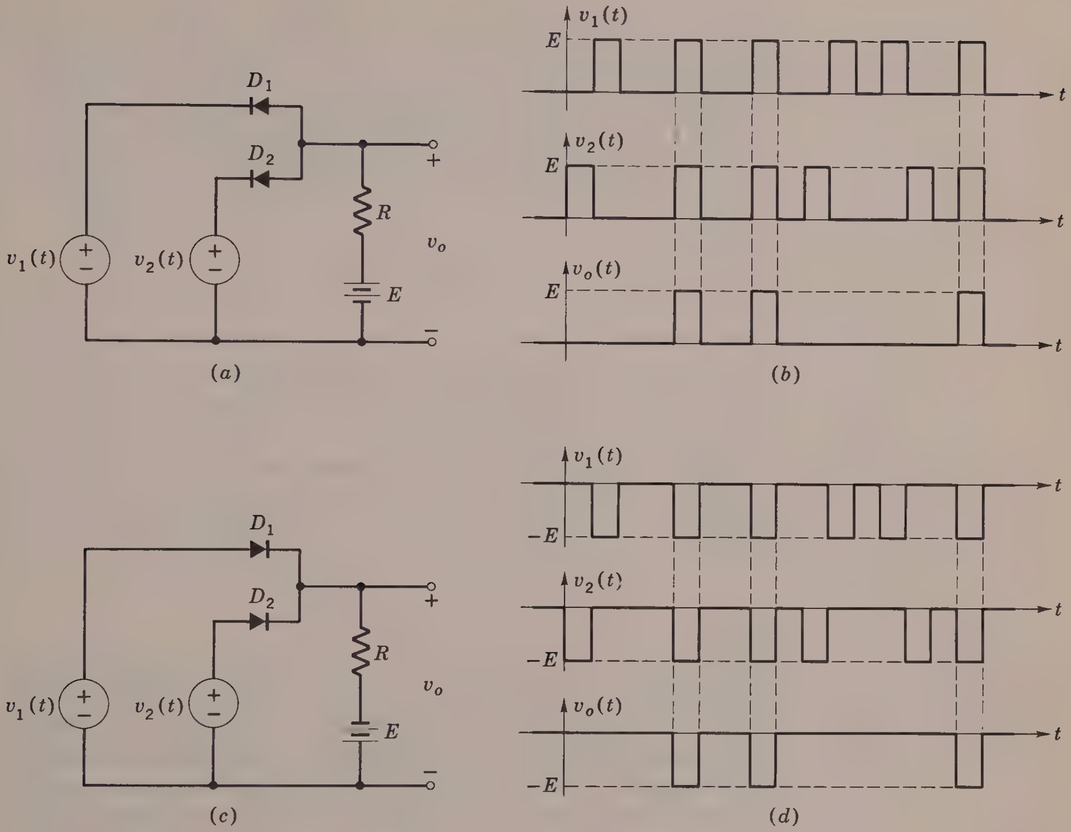


Fig. 10-17. Two circuits showing the complement of a positive logic AND gate is a negative logic AND gate.

Observe that this requirement can be satisfied by the *coincidence gate* shown in Fig. 10-13*a* with $v_{\text{eq}}(t) = E$ and with the switching signal amplitudes also set equal to E. In this case, our two switching levels are given by $E_1 = E$ and $E_0 = 0$. The resulting circuit is shown in Fig. 10-17*a* together with a set of typical input signals and the corresponding output signal as shown in Fig. 10-17*b*.

Notice that if we take the complementary network to the AND gate of Fig. 10-17*a* as shown in Fig. 10-17*c*,[1] we still obtain an AND gate provided we let the two levels be represented by $E_1 = -E$ and $E_0 = 0$. Notice that the waveforms shown in Fig. 10-17*d* are exactly the negative of the original waveforms in Fig. 10-17*b*. Since we are interested only in satisfying the truth table, the two AND gates in Fig. 10-17 are equally good realizations. In practice, however, sometimes one version is more convenient than the other. In order to distinguish between these two alternatives, however, it is customary to call the positive voltage version the *positive logic AND gate* and negative voltage version the *negative logic AND gate.* In fact, in view of the complementary-network theorem, it is clear that each of the remaining logic gates in Table 10-1 can be classified into *positive logic gates* and *negative logic gates,* depending on whether the signal waveforms are positive or negative. In fact, as a very useful corollary to the complementary-network theorem, we can state the following:

---

[1] This is drawn in accordance with the complementary-network theorem in Sec. 7-8. Observe that we can either reverse the polarities of the two input voltage sources or keep the polarities but change the sign of the original voltage waveforms. We have elected to choose the latter alternative as shown in Fig. 10-17*d*.

POSITIVE-NEGATIVE LOGIC NETWORK CONVERSION THEOREM

Any positive (negative) logic gate or combinations of positive (negative) logic gates can be converted into a negative (positive) logic gate or combinations of negative (positive) logic gates by replacing all nonlinear resistors by their complements,[1] and by reversing the terminals of all dc voltage and current sources. The polarities of all ac sources remain unchanged, but the signs of all waveforms are changed to the negative of the original waveforms.

10-6-2 SYNTHESIS OF OR GATES

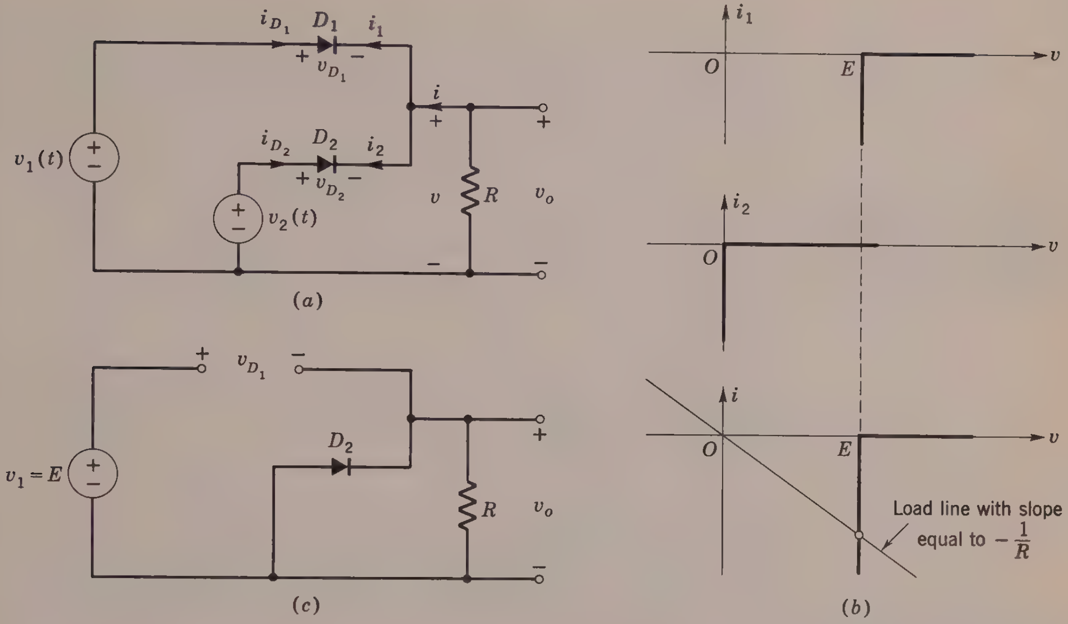
From the truth table in Table 10-1, an OR gate must give a nonzero output voltage $v_o = E_1$ when one or both of the input voltages are nonzero. The circuit shown in Fig. 10-18*a* is a simple realization of a two-input positive logic OR gate provided $E_1 = E > 0$ and $E_0 = 0$. In order to verify this realization, we only need to find the DP plot seen across the resistor as shown in Fig. 10-18*b* corresponding to a typical case $v_1(t) = E$ and $v_2(t) = 0$. Notice that in contrast with the case of a coincidence gate, the breakpoint of the DP plot is always equal to the *larger* of the breakpoint voltages of the component DP plots. This is why the breakpoint voltage of the DP plot of this circuit will never be zero as long as one input signal is not zero. The load-line construction shows that the circuit behaves exactly as a positive logic OR gate. A negative logic OR gate for this case is obtained simply by reversing the terminals of the diodes.

Logic gates using ideal diodes can often be analyzed by inspection using a few shortcuts which are peculiar to ideal diode networks. This is the *method of contradiction.* It is based on the observation that at any time, an ideal diode is either an open circuit or a short circuit. It is an open circuit if the voltage across the diode is negative, and a short circuit if the current flowing through the diode is positive.[2] In order to prove an assertion by the method of contradiction, we first assume that the assertion is not true, and

---

[1] Recall that the complements of ideal diodes are obtained by reversing their terminals; the complements of *p-n-p* (*n-p-n*) transistors are *n-p-n* (*p-n-p*) transistors; the complements of *p*-channel (*n*-channel) FETs are *n*-channel (*p*-channel) FETs.

[2] This rule is based on the usual reference direction and polarity as shown in Fig. 10-18*a*.



(a)

(c)

(b)

Fig. 10-18. The operation of a positive logic OR gate can be analyzed by either the graphical load-line method or the analytical method of contradiction.

going on this assumption, we try to establish a contradiction. If such a contradiction can indeed be established, then our original assumption is illegal, and hence the assertion is proved. The contradiction we seek consists of showing that the voltage across a supposedly open-circuited diode is positive. In the case of a supposedly short-circuited diode, we seek to show that the current flowing through it is negative. To illustrate this method, let us verify that the output voltage is equal to E when $v_1(t) = E$ and $v_2(t) = 0$. This assertion would require that diode D_1 be a short circuit. Hence, let us suppose the contrary and redraw the network as shown in Fig. 10-18*c* with D_1 replaced by an open circuit. This shows that $v_o = 0$ (since no current flows through the resistor R), and hence by KVL, $v_{D_1} = v_1 - v_o = E - 0 > 0$. But this contradicts the requirement that an ideal diode is an open circuit *only* if its terminal voltage is negative. This establishes the desired contradiction, and the assertion is proved.

With a little practice, it is possible to apply the method of contradiction mentally to a large class of ideal-diode logic circuits. It must be emphasized, however, that this method is applicable only for networks all of whose nonlinear elements are ideal diodes. If the ideal diodes are replaced by junction diodes, as is usually the case in practice, this method no longer applies, and

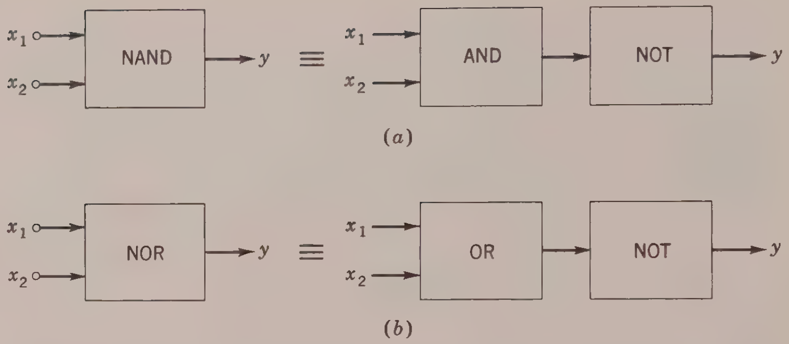


Fig. 10-19. A general method for realizing a NAND gate and a NOR gate consists of cascading an AND gate and an OR gate with a NOT gate.

other more general methods, such as the graphical load-line method, must be used.

10-6-3 SYNTHESIS OF NAND, NOR, AND NOT GATES

Let us now study how the NAND gate, the NOR gate, and the NOT gate in Table 10-1 may be synthesized. An examination of the truth tables in Table 10-1 shows that the output of the NAND gate is simply the opposite of the AND gate. Similarly, the output of the NOR gate is the exact opposite of the OR gate. Hence, these gates can be realized by connecting their outputs to some kind of an "inverting network" which converts an input voltage of level E_1 to an output voltage of level E_0, or vice versa. But this inverting network is exactly the NOT gate defined in Table 10-1. The above development shows that a NAND gate can be realized by cascading an AND gate with a NOT gate as shown in Fig. 10-19*a*. Similarly, a NOR gate can be realized by cascading an OR gate with a NOT gate as shown in Fig. 10-19*b*. These two realizations are valid provided the loading effects have been taken into consideration. The basic problem, therefore, consists of realizing a NOT gate.

If we let the two voltage levels be given by $E_1 = E$ volts and $E_0 = 0$ volt, then the output voltage of a NOT gate must be equal to E volts when the input voltage is zero, and vice versa. Recall that we have already encountered a network with this property, namely, the transistor unilateral switch shown in Fig. 10-5. This circuit is redrawn in Fig. 10-20*a*. The collector-to-emitter characteristics of the transistor corresponding to the two switching voltage levels $v_1 = 0$ and $v_1 = E$ are approximated by the two smooth curves as shown in Fig. 10-20*b*. From the intersection between these curves with the output load line, we found that when

$v_1(t) = 0$, $v_o(t) = E$, and when $v_1(t) = E$, $v_o(t) = 0$. The waveforms for $v_1(t)$ and $v_o(t)$ are, therefore, as shown in Fig. 10-20*c*. It must be pointed out, of course, that these are idealized waveforms. In actual transistors, the two idealized curves in Fig. 10-20*b* will not fall exactly on the set of axes, and hence the actual output voltage will vary somewhat from the ideal case shown in Fig. 10-20*c*.

If the resistor R_1 in Fig. 10-20*a* is chosen to be large enough so that the current flowing into the transistor base is not large, then we may connect the output terminals of an AND gate or an OR gate in place of the voltage source as shown in Fig. 10-20*d* and *e*, respectively. If the loading effects are negligible, the resulting networks will become a NAND gate or a NOR gate, respectively. Unfortunately, there will always be some loading effects, and this must be taken into consideration in any actual design.

If we recall that one of the characteristics of the FET is its almost infinite input resistance, then it is obvious that if we use the FET instead of the transistor, the loading effect might be negligible. An examination of the characteristic curves of a typical FET type 2NXX19 shows that in order to achieve the desired inverting effect, a bias battery E must be connected in series with the input voltage source as shown in Fig. 10-21*a*. Under this con-

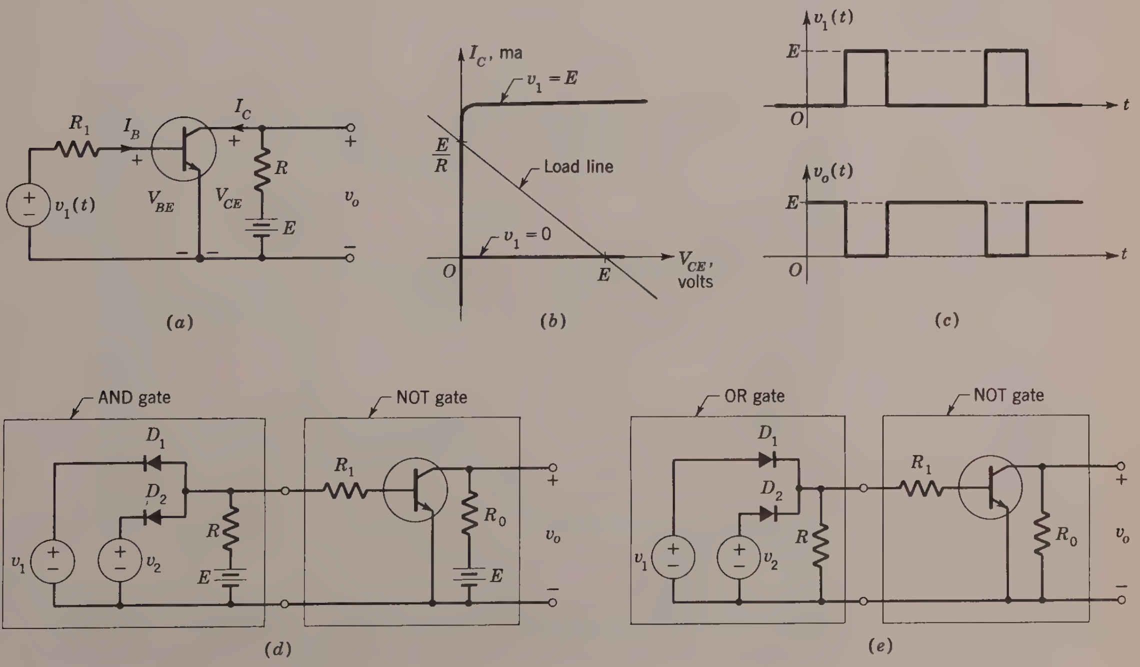


Fig. 10-20. A transistor NOT gate can be combined with an AND gate to form a NAND gate, or with an OR gate to form a NOR gate.

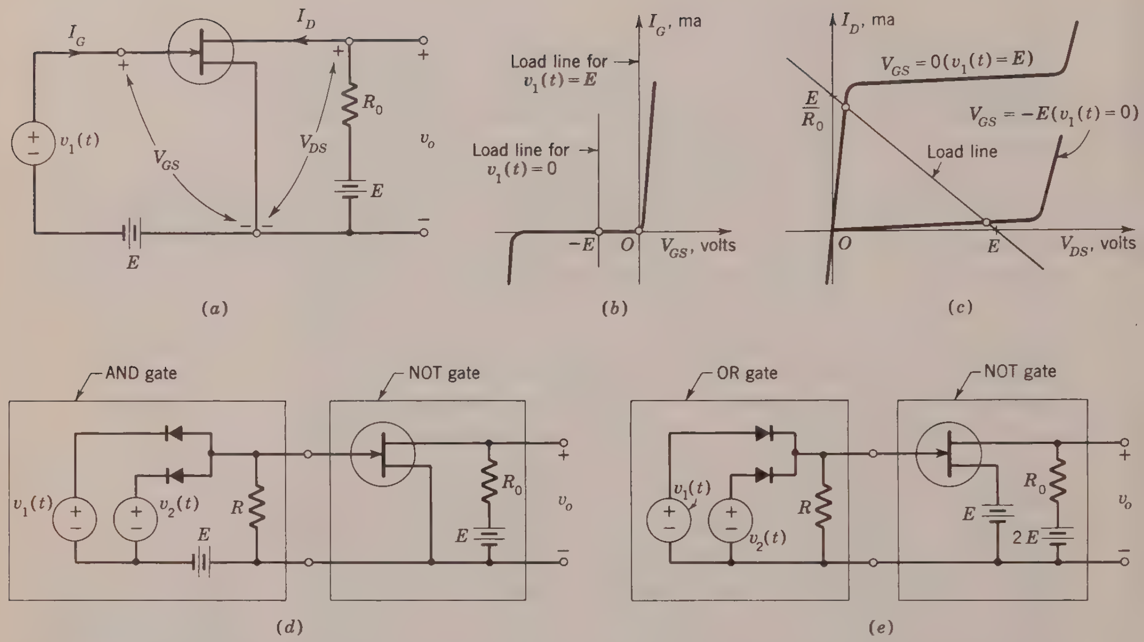


Fig. 10-21. An alternate realization of a NOT gate, NAND gate, and NOR gate using an FET to minimize loading effects.

dition, $I_G \approx 0$ for $v_1(t) = 0$ or $v_1(t) = E$, as shown by the load-line construction in Fig. 10-21*b*. Similarly, the output load line in Fig. 10-21*c* shows that $v_o(t) \approx E$ whenever $v_1(t) = 0$, and $v_o(t) \approx 0$ whenever $v_1(t) = E$. Since $I_G \approx 0$, we can feel confident this time in connecting the AND gate in place of the voltage source as shown in Fig. 10-21*d*, where the two batteries (one from the AND gate and one from the NOT gate) are replaced by one battery by means of the v-shift theorem. The result is, of course, a NAND gate. Similarly, a NOR gate is obtained by connecting the output terminals of the OR gate in place of the voltage source of the NOT gate as shown in Fig. 10-21*e*. Notice that this time we applied the v-shift theorem to move the battery E into the "source" terminal of the FET in order that all voltage sources would have a common ground. This particular feature is a highly desirable one in practice.

There are many other circuits which can be designed to function as various types of logic gates. Since their principles of operation are all analogous to those discussed in this chapter, we shall not labor further on this point. The reader is referred to the problems for more alternate circuit realizations.

Exercise 1: Generalize the definitions of the AND gate, OR gate, NAND gate, and NOR gate to include n input variables $x_1, x_2, \ldots, x_n$. Specify the truth table for each.

Exercise 2: The AND gate and the OR gate defined earlier in Chap. 3 [Eqs. (3-44) and (3-45)] have three additional requirements, that is, $i_1 = 0$, $i_2 = 0$, and $v_o = 0$ whenever $v_1(t) < 0$ or $v_2(t) < 0$. (*a*) Show why the first two requirements $i_1 = 0$ and $i_2 = 0$ are desirable, even though they cannot be exactly attained in practice. (*b*) Show that the third requirement can be approximately satisfied in practice by connecting a junction diode at appropriate locations.

Exercise 3: Without using the complementary-network theorem, verify that the circuit shown in Fig. 10-17*c* is indeed a negative logic AND gate.

Exercise 4: Verify the properties of the positive logic AND gate in Fig. 10-17*a* by the method of contradiction.

Exercise 5: Verify the properties of a negative logic OR gate using ideal diodes by the method of contradiction.

Exercise 6: (*a*) Show that if we apply the positive voltage waveforms $v_1(t)$ and $v_2(t)$ of Fig. 10-17*b* to the negative logic AND gate in Fig. 10-17*c*, the result is a positive logic OR gate. (*b*) Show that a negative logic OR gate is obtained if we apply the negative voltages $v_1(t)$ and $v_2(t)$ of Fig. 10-17*d* to the positive logic AND gate in Fig. 10-17*a*. (*c*) Show that in both cases (*a*) and (*b*), the battery E can be replaced by a short circuit.

Exercise 7: The circuits shown in Figs. 10-20 and 10-21 are all positive logic gates. Draw the corresponding negative logic gates and explain their principles of operation.

Exercise 8: Design a NOT gate using an *n-p-n* transistor type 2NXX17B. Assume that $E_1 = 10$ volts and $E_2 = 0$. Discuss any disparity between the output waveform and the ideal waveforms shown in Fig. 10-20*c*.

10-7 SUMMARY

Synthesis of electronic switches The basic core of any electronic switch consists of one of the following basic components:

1. A single-controlled switch which can be turned on or off by a single switching signal.
2. A multicontrolled switch which can be turned on or off by a combination of switching signals.

Each switch is unidirectional if it works only for single-polarity input signals. Otherwise, it is bidirectional.

Classification of electronic switches

1. Series switch
2. Shunt switch

Each switch may be a latching type or a nonlatching type.

Switching sensitivity This is defined as the ratio between the maximum amplitude of the input signal and the minimum amplitude of the required switching signal in order to ensure satisfactory operation. The smaller the amplitude of the switching signal needed to turn on or off a switch, the higher is its switching sensitivity.

Some applications of electronic switches

1. Choppers (e.g., Cowan modulator)
2. Amplitude modulators (e.g., Ring modulator)
3. Coincidence gate
4. Transmission gate

Logic building blocks These include both positive and negative logic blocks. The five most basic components are as follows:

1. AND gate
2. OR gate
3. NAND gate
4. NOR gate
5. NOT gate

PROBLEMS

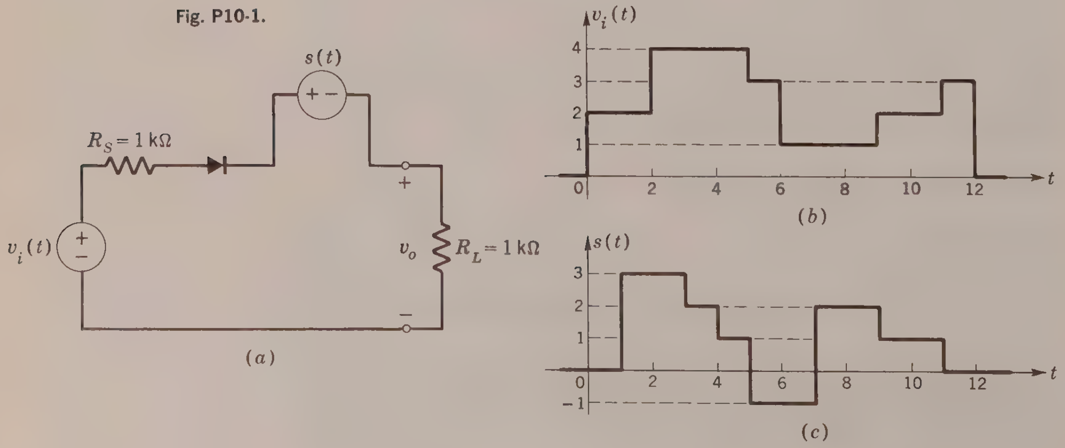
10-1 Consider the series switching circuit configuration shown in Fig. P10-1*a*, where the input voltage $v_i(t)$ and the switching signal $s(t)$ are given in Fig. P10-1*b* and *c*, respectively.

(a) Find the waveform of the output voltage $v_o(t)$.

(b) Using the results in (*a*), formulate a simple set of rules for determining the output voltage by inspection.

(c) If the ideal diode is replaced by a junction diode type 1NXX2, find the resulting waveform of the output voltage. Compare this with (*a*).

Fig. P10-1.



10-2 Repeat Prob. 10-1 for the shunt switch configuration, that is, with the diode-switching source combination connected across the load resistor.

10-3 Using *n-p-n* transistor type 2NXX17B, design the single-controlled transistor switching circuit shown in Fig. 10-5*h* with $R_s = 150\ \Omega$, $R_L = 300\ \Omega$, $v_i = 8 + 5 \cos \omega t$, and $E < 0.4$ volt [E is the amplitude of $v_s(t)$].

10-4 Using a *p-n-p* transistor type 2NXX18B, design an electronic switch with $R_s = 200\ \Omega$, $R_L = 300\ \Omega$, and $v_i = -5(1 + \sin \omega t)$.

10-5 Consider the shunt transistor switching configuration shown in Fig. P10-5*a* where $v_i(t)$ and $v_s(t)$ are shown in Fig. P10-5*b* and *c*, respectively.
(a) Find $v_o(t)$.
(b) If R_B is increased to 250 Ω, find the resulting $v_o(t)$. What can you conclude?
(c) Find the minimum value of R_s in order for $v_o(t)$ in (*b*) to come out right.
(d) Find the switching sensitivity.

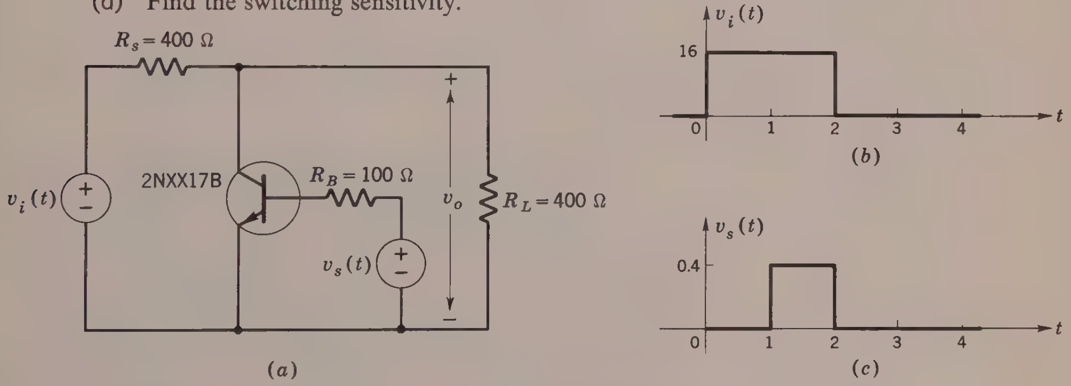


Fig. P10-5.

10-6 Consider the series latching-switch circuit configuration shown in Fig. P10-6*a*, where $v_i(t)$ and $i_s(t)$ are as shown in Fig. P10-6*b* and *c*, respectively.
(a) Find $v_o(t)$ and the switching sensitivity.
(b) Repeat (*a*) for the case where $R_L = 100\ \Omega$.

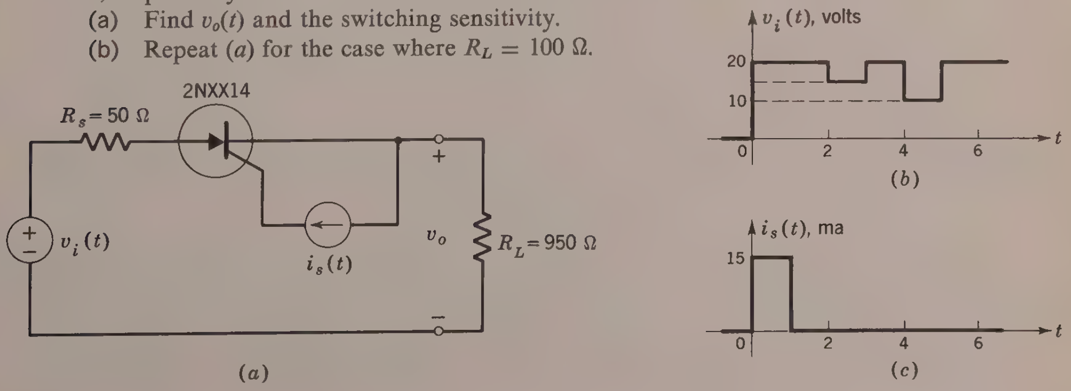


Fig. P10-6.

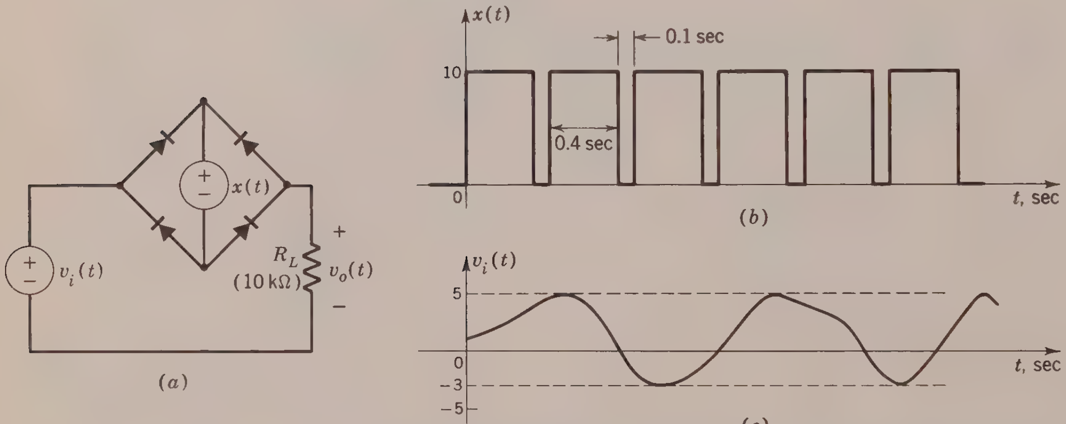
10-7 Consider the bidirectional electronic switch shown in Fig. P10-7*a*.

(a) Find the TC plot between v_o and v_i when $x(t) = 0$, 5, and 10 volts.

(b) If the maximum absolute value of $v_i(t)$ is 8 volts, find the minimum value of $x(t)$ in order for the circuit to function as a chopper.

(c) Find the switching sensitivity of the chopper.

(d) If $x(t)$ and $v_i(t)$ are given in Fig. P10-7*b* and *c*, respectively, find $v_o(t)$.



(*a*) (*b*) (*c*)

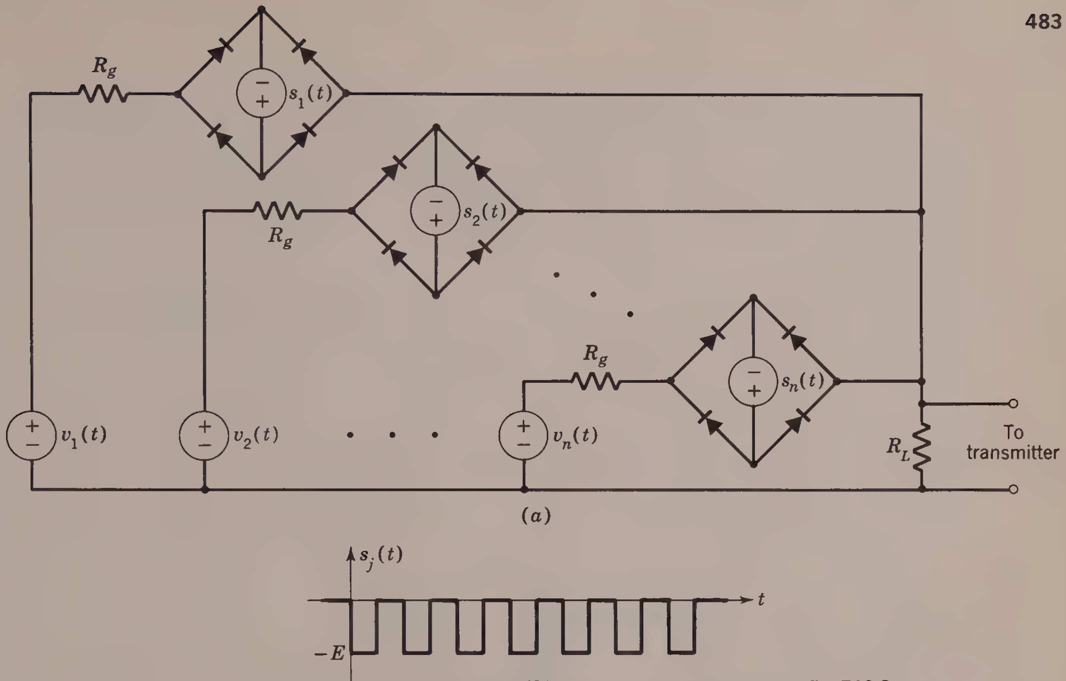
Fig. P10-7.

10-8 It is often desirable to transmit several unrelated signals through the same transmission channel in order to save weight and cost. For example, in a certain satellite, only one transmitter was used to transmit signals from various sensing instruments for measurements of pressure, temperature, radiation intensity, cosmic dust, etc. In order to accomplish this task using one transmitter, a communication system was designed using the time-share multiplex principle, wherein the transmitter takes samples of each signal one at a time and repeats this sampling in quick succession. If the interval between sampling is short enough, then by the well-known Nyquist theorem it is possible to design a receiver which will reproduce the desired set of signals satisfactorily. The heart of this communications system is nothing but a suitably designed series of choppers connected as shown in Fig. P10-8*a*. A typical switching signal is shown in Fig. P10-8*b*.

(a) If the magnitude of all signals $v_j(t)$ does not exceed $v_{\max}$ for all time, what must be the relationship between the values of $V_{\max}$, R_g, R_L, and E?

(b) If the signals $v_1(t), v_2(t), \ldots, v_n(t)$ are to be sampled in the order $1, 2, 3, \ldots, n$, sketch the required waveforms of the switching signals $s_1(t), s_2(t), \ldots, s_n(t)$.

10-9 The circuit shown in Fig. P10-9*a* is used to provide signals to an amplifier from a semiconductor strain gage. The characteristics



(a)

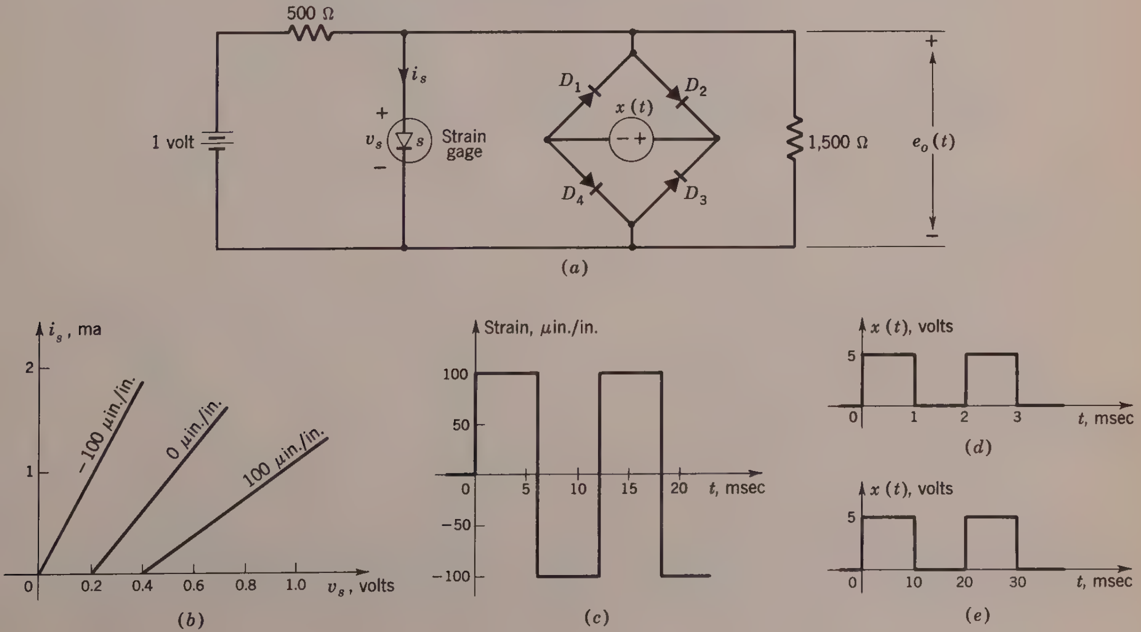
(b)

Fig. P10-8.

for the strain gage are shown in Fig. P10-9*b*. For test purposes, the gage is attached to a beam and subjected to a strain as shown in Fig. P10-9*c*.

(a) Plot $e_o(t)$ when $x(t)$ is given by Fig. P10-9*d*.
(b) Plot $e_o(t)$ when $x(t)$ is given by Fig. P10-9*e*.

Fig. P10-9.



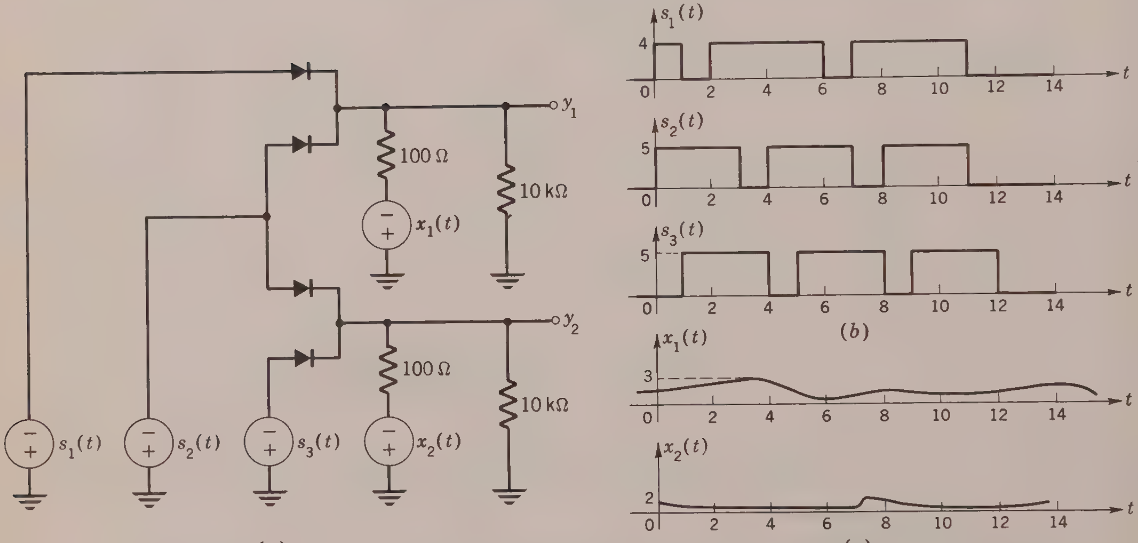
(a)

(b) (c) (d) (e)

10-10 Consider the two-input, two-output transmission gate shown in Fig. P10-10*a*. Let the controlling signals be as shown in Fig. P10-10*b* and the input signals be as shown in Fig. P10-10*c*.

(a) Find the output signals.

(b) Find the maximum allowable values for $x_1(t)$ and $x_2(t)$.



(*a*)

(*b*)

(*c*)

Fig. P10-10.

10-11 Design a matrix transmission gate to implement the switching combinations listed in the following table. It is desired that the output signal $y_j(t)$ be equal to the input signal $x_j(t)$ whenever the switching signals $s_j(t)$ assume the value E according to the combinations listed in the table.

	s_1	s_2	s_3	s_4	s_5	s_6	
x_1		E	E			E	y_1
x_2	E	E			E	E	y_2
x_3	E			E	E		y_3
x_4			E	E			y_4
x_5	E					E	y_5
x_6	E	E	E		E		y_6
x_7			E	E		E	y_7
x_8		E			E		y_8
x_9				E	E	E	y_9
x_{10}	E		E	E			y_{10}

10-12 By an appropriate circuit arrangement of thermistors with different *v-i* curves at a given temperature (or identical thermistors, each heated to a different temperature), it is possible to design a wide variety of latching-combination switches. For this problem, consider the circuit shown in Fig. P10-12. Assume that the v-i curves of R_1, R_2, and R_3 are given, respectively, by the NTC thermistor (type 1NXX12) curves corresponding to $T = 75$, 50, and 25°F.

(a) With the switch S_1 closed, find the DP plot across terminals a-b.

(b) Find the operating point of the circuit when S_3 is closed after S_1 is closed.

(c) Find the operating point of the circuit when S_3 is closed before S_1 is closed.

(d) What is the effect of closing S_3 after both S_1 and S_2 are closed?

(e) Describe the effects of the sequence in the closing of S_1, S_2, and S_3 on the circuit performance. Can you suggest some practical applications?

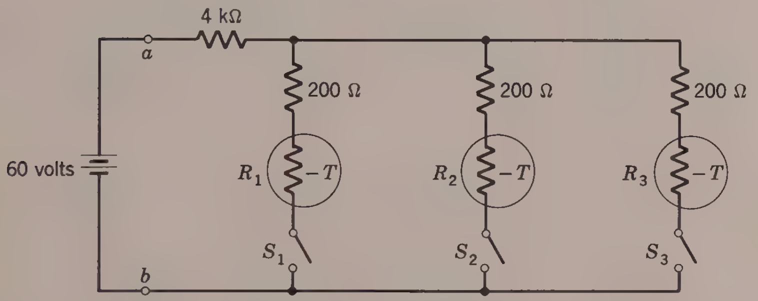


Fig. P10-12.

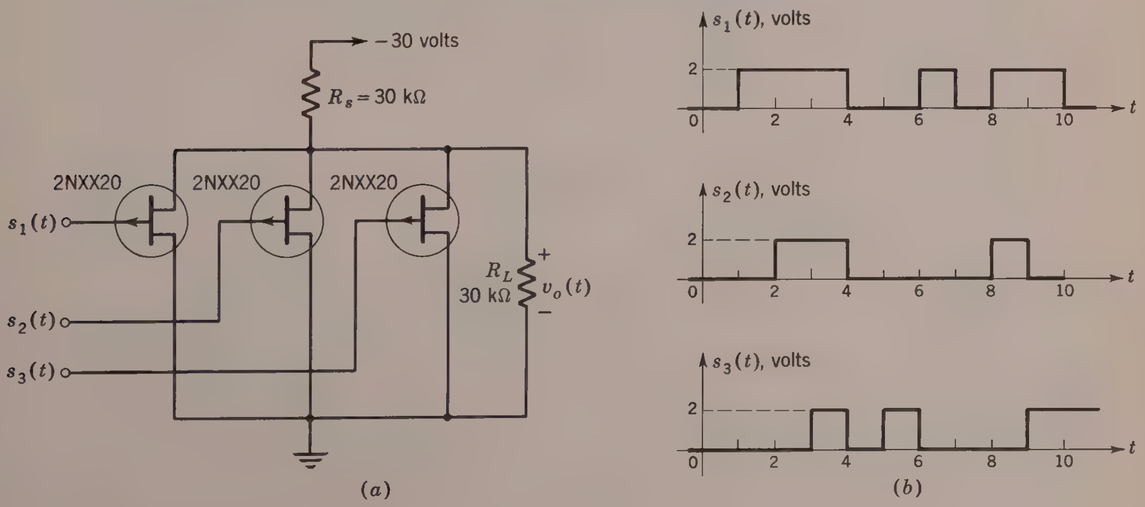
10-13 Consider the FET circuit shown in Fig. P10-13a.

(a) If the signals $s_1(t)$, $s_2(t)$, and $s_3(t)$ are as shown in Fig. P10-13b, find $v_o(t)$.

(b) What kind of logic functions does this circuit perform?

(c) Can you find an alternate realization using transistors?

Fig. P10-13.



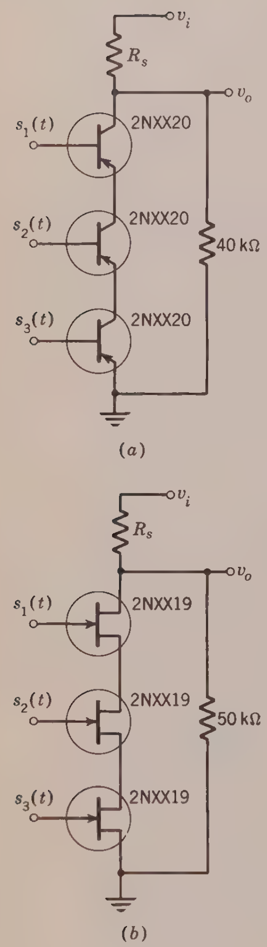


Fig. P10-14.

10-14 Consider the string of series transistor and FET circuits shown in Fig. P10-14*a* and *b*.

(a) Analyze these circuits and find what kind of logic functions they perform.

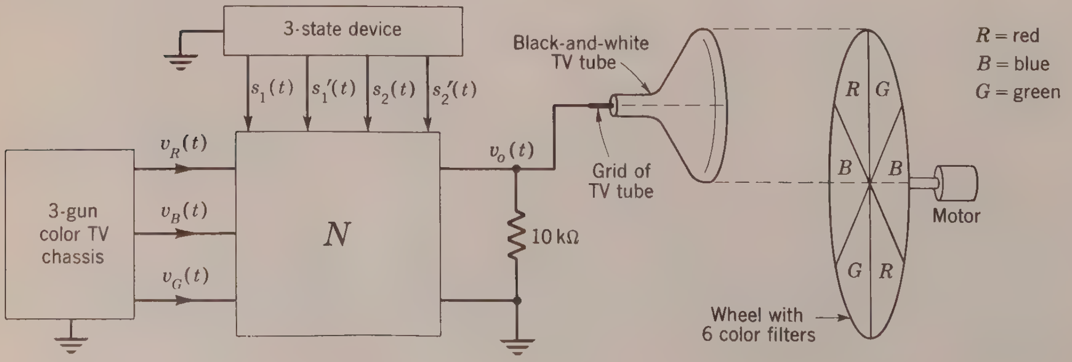
(b) Find the remaining circuits in order that both positive and negative logic functions can be performed.

10-15 You happen to have an old color TV chassis designed to be used with a commercial three-gun color tube. Unfortunately, you do not have that tube and cannot afford one. Rather than display the three color components simultaneously (as in a three-gun tube) you choose to display them sequentially. Assume that you have solved all your problems except how to gate the three video signals (each of whose amplitudes ranges from 0 to 80 volts) sequentially to the grid of the TV tube, such that the appropriate color signal is present at the grid when the corresponding color filter is passing the face of the TV tube. See Fig. P10-15. Your three-state device produces the following outputs versus color filters at the tube face (assume that the voltage level $E = 40$ volts):

Color filters at tube face	Output voltage			
	s_1	s_1'	s_2	s_2'
Red	0	E	0	E
Blue	E	0	0	E
Green	E	0	E	0

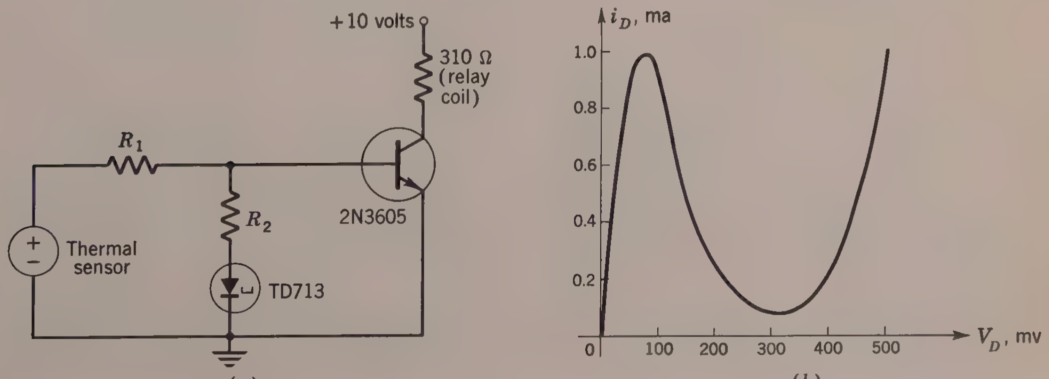
Using no more than nine ideal diodes and three linear resistors, design the black box N choosing the resistors so as to maximize the signal $v_o(t)$ without clipping effects. You may represent the input signals $v_R(t)$, $v_B(t)$, $v_G(t)$, $s_1(t)$, $s_1'(t)$, $s_2(t)$, and $s_2'(t)$ by ideal voltage sources. You may also assume that the grid of the TV tube does not draw any current.

Fig. P10-15.

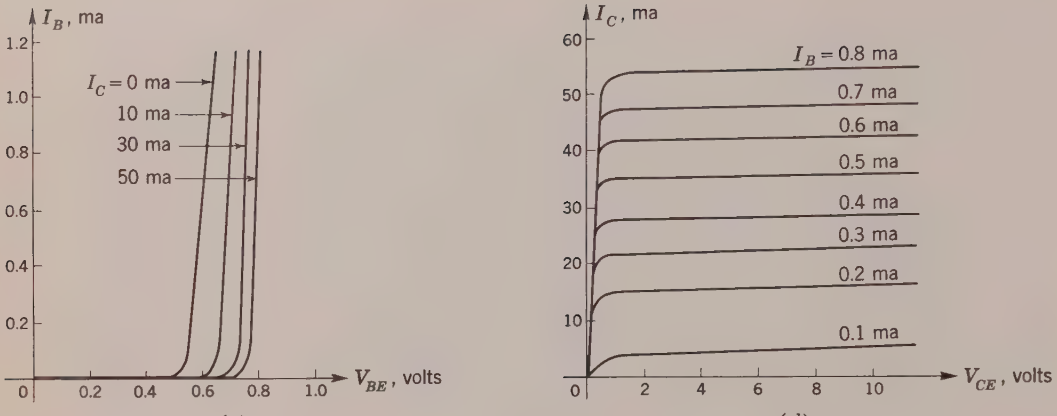


10-16 The advanced-engineering department has proposed a fire-warning system and fire-extinguishing actuator for an aircraft jet engine. A fire-detecting thermovoltaic sensor puts out a 1.1-volt direct current when the temperature in the engine nozzle exceeds 600° C. However, the sensor can produce a current of only 1.5 ma. The relay that actuates the fire indicator and fire-extinguishing system requires 15 ma through its coil before the contacts close. In order to prevent spurious closing of the relay, a preliminary engineering design suggests a threshold detector that will actuate the relay when the sensor output is 1.0 volt or more. Such a detector is shown in Fig. P10-16*a*. The circuit is to be designed to actuate the relay when the sensor voltage is equal to or greater than 1.0 volt. Your job will be to choose R_1 and R_2 so that the circuit will operate properly. The V_D-I_D curve of the tunnel diode chosen for this circuit is shown in Fig. P10-16*b*. The input- and output-characteristic curves of the transistor are shown in Fig. P10-16*c* and *d*, respectively. The following information has been provided by the preliminary design engineer: (1) The sum of R_1 and R_2 must be greater than the maximum negative resistance of the tunnel diode. (2) The ratio of R_1 to R_2 is selected to provide enough base current in the transistor to support the collector current necessary to close the relay. (3) The maximum ratings on the tunnel diode are maximum forward current = 5 ma, maximum reverse current = 10 ma. (4) The maximum ratings on the transistor are maximum collector-to-emitter voltage = 14 volts, maximum emitter-to-base voltage = 5 volts, maximum collector current = 200 ma, maximum collector dissipation at 25° C = 200 mw. (5) Once the relay closes it is not necessary for it to open until the sensor voltage reaches zero. (6) The relay coil current should remain at relatively low values for sensor voltages below 1 volt, but the coil current should jump to 15 ma or slightly greater as soon as the sensor voltage reaches 1.0 volt (threshold detecting action). With the above information select R_1 and R_2 so that the system will operate as desired.

Fig. P10-16.



(*a*) (*b*)



(c) (d)

Fig. P10-16. *(Continued)*

10-17 An electronic relay can be considered as a two-port black box with port 1 used as the relay terminals and with port 2 connected to a switching signal $s(t)$, where $s(t)$ can be either a voltage or a current waveform.

(a) A "normally opened" electronic relay is characterized by

$$v_1 = 0 \qquad \text{whenever } s(t) \geq k$$

$$i_1 = 0 \qquad \text{whenever } s(t) < k$$

where k is some prescribed threshold level above which the relay closes. Assuming that $s(t)$ is a voltage source, find a circuit realization of this relay using two linear controlled sources, two ideal diodes, and an appropriate nonlinear resistor.

(b) Repeat (*a*) for the case where $s(t)$ is a current source.

(c) A "normally closed" electronic relay is characterized by

$$v_1 = 0 \qquad \text{whenever } s(t) \leq k$$

$$i_1 = 0 \qquad \text{whenever } s(t) > k$$

where k is some prescribed threshold level above which the relay opens. Assuming that $s(t)$ is a voltage source, find a circuit realization of this relay using two linear controlled sources, two ideal diodes, and an appropriate nonlinear resistor.

(d) Repeat (*c*) for the case where $s(t)$ is a current source.

NOTE: In all cases, assume the signal magnitude across port 1 of the relay is bounded by a prescribed range; namely, $E_{\min} < v_1 < E_{\max}$ and $I_{\min} < i_1 < I_{\max}$. Assume also that the switching signal $s(t)$ is a continuous function of time.

11 SYNTHESIS OF RESISTIVE NONLINEAR NETWORK MODELS

11-1 WHAT IS A NETWORK MODEL?

Most of the analysis and synthesis methods presented in the preceding chapters are valid only for networks containing two-terminal nonlinear resistors and independent sources. At present, there are no general methods available for handling arbitrary networks containing nonlinear controlled resistors and n-terminal resistors ($n > 2$). The basic difficulty lies in the fact that these nonlinear network elements are characterized by families of curves, rather than just one curve, as in the case of a two-terminal resistor. In some special cases, it may be possible to represent these curves *approximately* by mathematical equations. Since these equations are at best approximations, they are said to be mathematical *models* rather than mathematical *characterizations.*

For example, by studying the physical phenomena inside a transistor, Ebers and Moll succeeded in deriving the following two equations which now bear their names:[1]

$$I_E = I_{ES}(e^{qV_{EB}/kT} - 1) - \alpha_R I_{CS}(e^{qV_{CB}/kT} - 1) \tag{11-1}$$

$$I_C = -\alpha_F I_{ES}(e^{qV_{EB}/kT} - 1) + I_{CS}(e^{qV_{CB}/kT} - 1) \tag{11-2}$$

where I_E and I_C are the emitter and collector currents; V_{EB} and V_{CB} are the emitter-to-base and collector-to-base voltages; q and k are physical constants (representing electron charge and Boltzmann's constant); T is the absolute temperature; and I_{ES}, I_{CS}, α_R, and α_F are parameters which depend on the construction of the individual transistor and must be measured experimentally for *each* transistor. The Ebers-Moll equations are a very good mathematical model of a transistor. For example, the characteristics of a typical *n-p-n* transistor (Fig. 11-1*a*) are shown by the solid curves in Fig.

---

[1] The reader is encouraged to read the classic paper, J. J. Ebers and J. L. Moll, Large Signal Behavior of Junction Transistors, *Proc. IRE,* vol. 42, pp. 1761–1772, December, 1954.

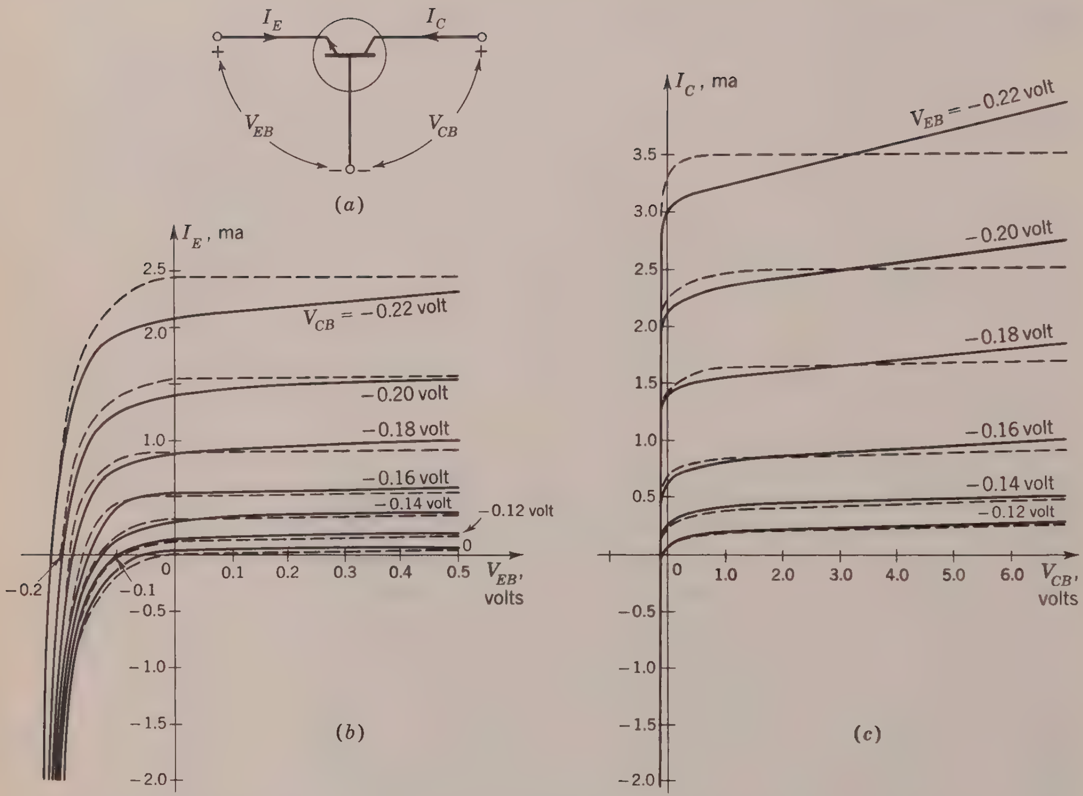


Fig. 11-1. A comparison between the experimentally measured input- and output-characteristic curves of a typical common-base *n-p-n* transistor (solid curves) and those predicted by the Ebers-Moll equations (dotted curves).

11-1*b* and *c*. The corresponding characteristics predicted by the Ebers-Moll equations (with the parameters I_{ES}, I_{CS}, α_R, and α_F measured from the same transistor) are shown by the dotted curves which are superimposed on top of the experimentally measured curves. Notice that the approximation is very good indeed. The only significant feature of the actual transistor which was not predicted is the slope of the curves. The curves obtained from the Ebers-Moll equations are almost horizontal, whereas the actual curves show that the slope increases as the curves fan out.

If a network contains only transistors and diodes for which accurate mathematical models are available, then we need only write the equations of motion of the network as was done in Chap. 4 and use *numerical techniques* to find the solution. Unfortunately, most practical devices do not have good mathematical models. In such cases, we must find some alternate methods to analyze networks containing these elements.

The above difficulty can be overcome if it is possible to replace each controlled resistor and each n-terminal resistor in a given network by an *equivalent* black box† containing only two-terminal nonlinear resistors, linear elements (such as ideal transformers, gyrators, rotators, etc.), and sources (independent and dependent sources). This results in a network whose only nonlinear elements are two-terminal resistors. This class of network can be analyzed by the iterative piecewise-linear method to be given in the next chapter.

Unfortunately, we do not know of a general method for synthesizing the equivalent black box described above for arbitrary controlled resistors and arbitrary n-terminal resistors. The next-best thing that we can do is to find a black box which *approximately* represents these elements. We shall refer to such a black box as the network model, or simply the model, of the corresponding controlled resistor or n-terminal resistor.

As an example, suppose the hypothetical controlled resistor shown in Fig. 11-2*a* is characterized by the family of v-i curves shown in Fig. 11-2*b*.‡ In order to be able to analyze a network containing this controlled resistor, we must first synthesize an equivalent network containing only two-terminal nonlinear resistors, linear elements, and sources. Since no method is currently known for realizing the curves of Fig. 11-2*b*, we are forced to approximate the characteristics of this controlled resistor by the curves shown in Fig. 11-2*c*.

This family of curves can be realized by the simple network shown in Fig. 11-2*d*, where the v_1-i_1 curve of the two-terminal resistor is shown in Fig. 11-2*e*. Hence, the network shown in Fig. 11-2*d* is a model of the controlled resistor. Notice that this is a network model but not a mathematical model because it is not represented by mathematical equations. Since this model does not give an exact representation of the curves in Fig. 11-2*b*, the solutions we obtain based on this model will only approximate the correct solution. Obviously, there is no sense making models unless the discrepancy between the resulting solution and the correct solution is negligible. Hence, a network model may be interpreted as an approximate equivalent circuit.

Exercise 1: What is the difference between a mathematical model, a network model, and an equivalent circuit?

Exercise 2: What are the relative advantages of a network model and a mathematical model?

---

† The definition of equivalent networks in Chap. 7 requires that an equivalent black box must be indistinguishable externally from the corresponding controlled resistor or n-terminal resistor that it represents.

‡ A controlled source is used here in place of an independent source because (as will be shown in the following sections) the controlling variable usually represents some voltage or current in the network model of an n-terminal device.

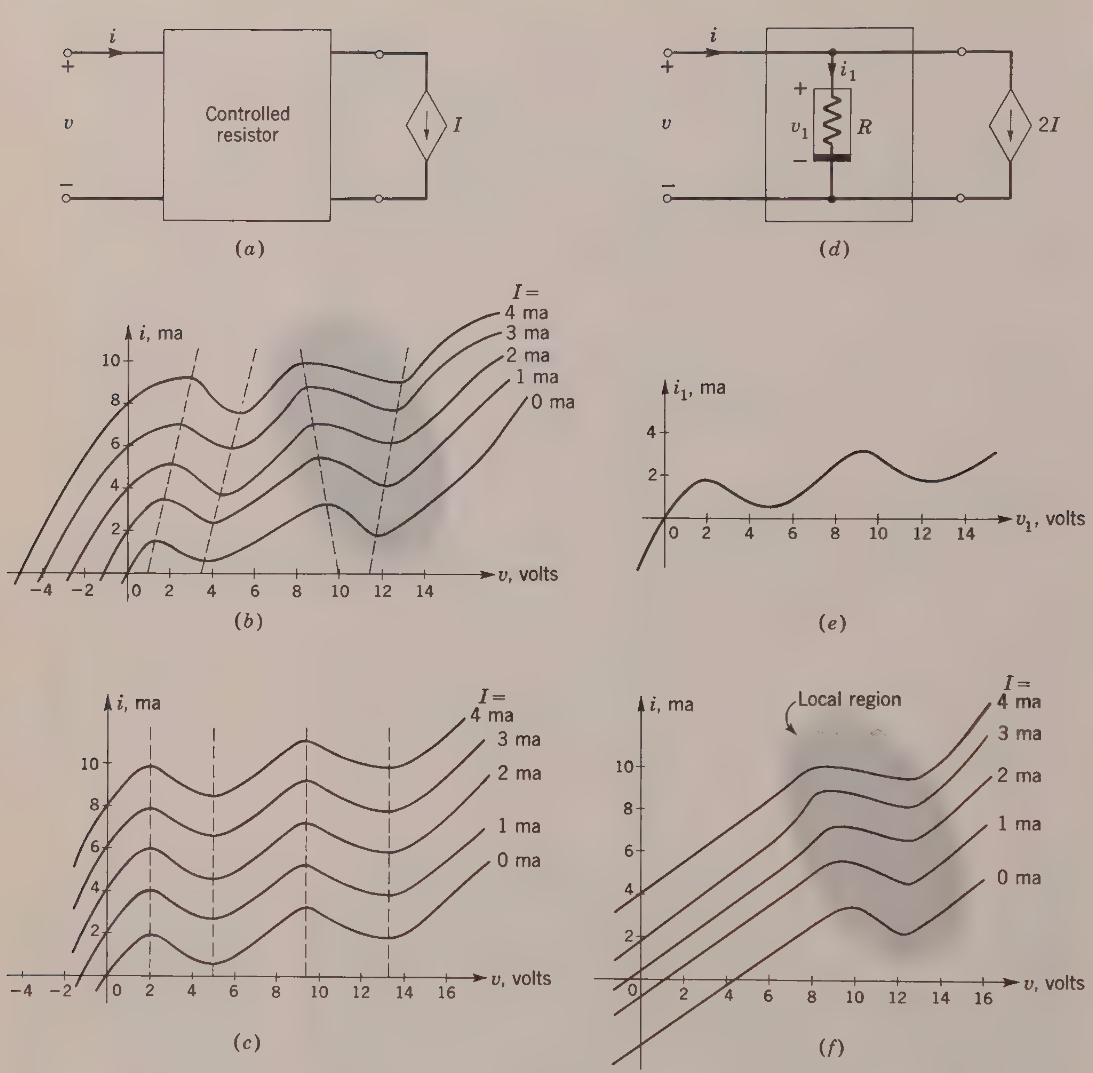


Fig. 11-2. An example illustrating the network model of a hypothetical nonlinear controlled resistor.

11-2 PRINCIPLES OF MODEL MAKING

The basic principles of model making are identical with the principles of network synthesis presented in Chap. 8. In fact, model making and network synthesis differ only in the ultimate objectives. In model making, we have no intention to build the synthesized network, and hence the availability of the components to be used is of no concern to us. For example, in the model shown in Fig. 11-2*d*, there is no need to replace the resistor R by a combination of concave resistors, convex resistors, and a negative resistor because we shall never build the network anyway. The purpose of this model is simply to convert the controlled resistor into a network containing two-terminal resistors and sources, so that the resulting network can be analyzed by the general methods to be presented in the next chapter. Hence, we are free to use controlled sources, ideal diodes, and *any* two-terminal nonlinear resistors as components in the network models.

Given a controlled resistor or an *n*-terminal resistor there are generally many distinct models that can be found to approximate the resistor's characteristics. Clearly, different models would yield different degrees of approximation. In general, the more complicated the model, the more accurate is the approximation. However, in constructing the model for a given resistor, we must always bear in mind that, ultimately, we have to analyze the network whose controlled resistors and *n*-terminal resistors are replaced by their models. We do not want to complicate the analysis by using a complicated model if we can obtain a realistic solution with a simpler model. After all, as engineers, we are only after a realistic and practical solution. Very seldom do we need to know the exact answer. Hence, it would be foolish to try to use a model more complicated than necessary. In this respect, a certain model may be perfectly satisfactory when used in one network but may be completely unsatisfactory when used with another network. In other words, there is no such thing as the "best" model for a given controlled resistor or *n*-terminal resistor unless we are referring to the model with respect to a particular network containing the resistor, and the criteria for evaluating the model are specified.

A question that arises in connection with the choice of models is, how do we choose the simplest from among a number of models that would yield realistic solutions? It is clearly impossible to answer this question in general since much depends on a person's engineering judgment and previous experience working on networks containing similar elements. It is reasonable, however, to start with a fairly simple, if not the simplest, model; if the solution turns out to be unrealistic, then we repeat the problem with a better model. It must be emphasized again that should a certain model in a certain problem result in an unrealistic solution, it does not mean that the method or the theory is wrong. It means only that a better model should be chosen for that particular problem.

The network model shown in Fig. 11-2*d* is a fairly good representation of the controlled resistor throughout the entire v-i plane. In practice, however, the operating point of the resistor usually lies within a small region of the v-i plane. For example, the operating point may lie only within the shaded local region shown in Fig. 11-2*b*. In this case there is no need to find a model which represents the controlled resistor over the entire v-i plane. In fact, a model representing the v-i curves shown in Fig. 11-2*f* is perfectly acceptable even though the curves outside the local region are radically different from the actual curves of Fig. 11-2*b*. Indeed, by concentrating over a certain region of the v-i plane where the operating point is known to be located, we can usually synthesize a more accurate model than one that covers the entire plane. However, outside the local region the model is unsatisfactory. In order to differentiate between these two types of models, we shall refer to any model which covers the entire v-i plane as a *global model*. Similarly, we shall refer to any model which covers only a region of the v-i plane as a *local model*. Sometimes, the pertinent region of the v-i plane may be small enough that the family of curves within this region are parallel straight lines. In this case the local model can be represented by a linear network with batteries. (See Fig. 8-4*a*.) We shall refer to this as a *local linear model.*[1]

11-3 SYNTHESIS OF GLOBAL MODELS FOR NONLINEAR CONTROLLED RESISTORS

We have already shown in Chap. 2 that a controlled resistor is a two-terminal resistor having a v-i curve that changes with respect to one or more controlling parameters. The controlling parameter may be some external voltage or current or it may be some nonelectrical variable such as temperature, light intensity, or pressure. Since most practical, controlled resistors have only one controlling parameter x, we shall synthesize models only for this type of controlled resistor. For convenience, we shall assume that the family of v-i curves of the controlled resistor has been approximated by piecewise-linear segments. We shall further assume that these curves satisfy the following two properties:

1. All v-i curves are parallel with respect to some straight line α through the origin of the v-i plane. By this we mean that the corresponding segments of the v-i curves are parallel to one another, and the corresponding breakpoints lie along a straight line parallel to α.

---

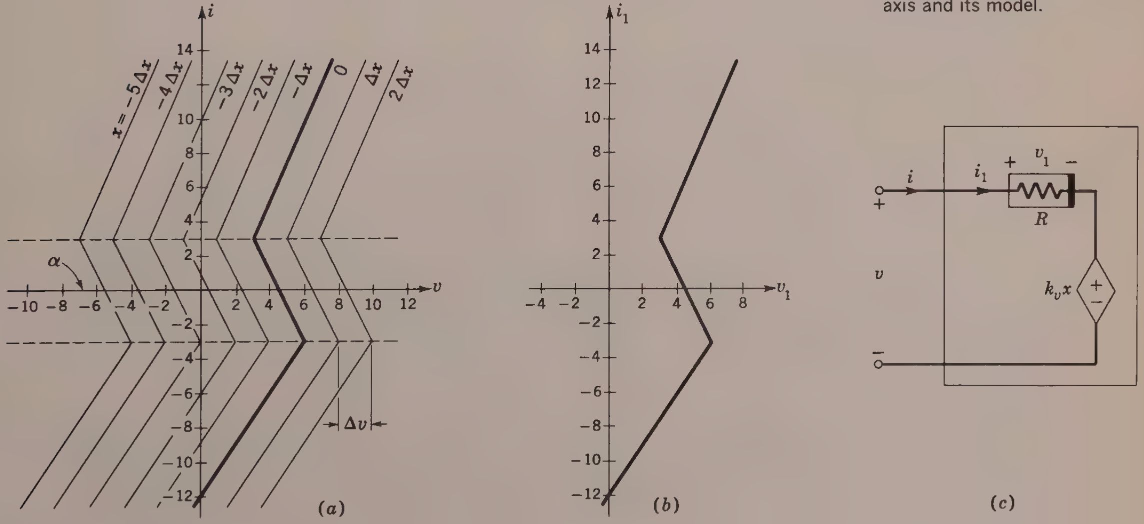
[1] In some practical circuits such as small-signal amplifiers, the signal is so small that a local linear model can always be constructed about any operating point. Such a model is called an *incremental linear model* in order to emphasize the fact that it is valid only for analyzing small-signal circuits. In general, the parameters associated with the incremental linear model of a given device would vary with the location of the operating points.

2. All v-i curves are uniformly spaced; that is, if the controlling variable x changes by some fixed increment Δx (say, $x = \ldots, -3\,\Delta x, -2\,\Delta x, -\Delta x, 0, \Delta x, 2\,\Delta x, 3\,\Delta x, \ldots, n\,\Delta x$), then the current intercept also changes by some fixed increment Δi, and the voltage intercept changes by some fixed increment Δv.

Examples of some v-i curves with the above properties are shown in Figs. 11-3*a*, 11-4*a*, and 11-5*a*. Observe that in view of property 1, this family of curves can be generated by translating the v-i curve corresponding to $x = 0$ along the straight line α. Accordingly, we shall call this curve the *v-i curve generator.* Observe also that in view of property 2, it is possible to use a linear controlled source to translate the v-i curves by uniform increments. Let us consider some special cases:

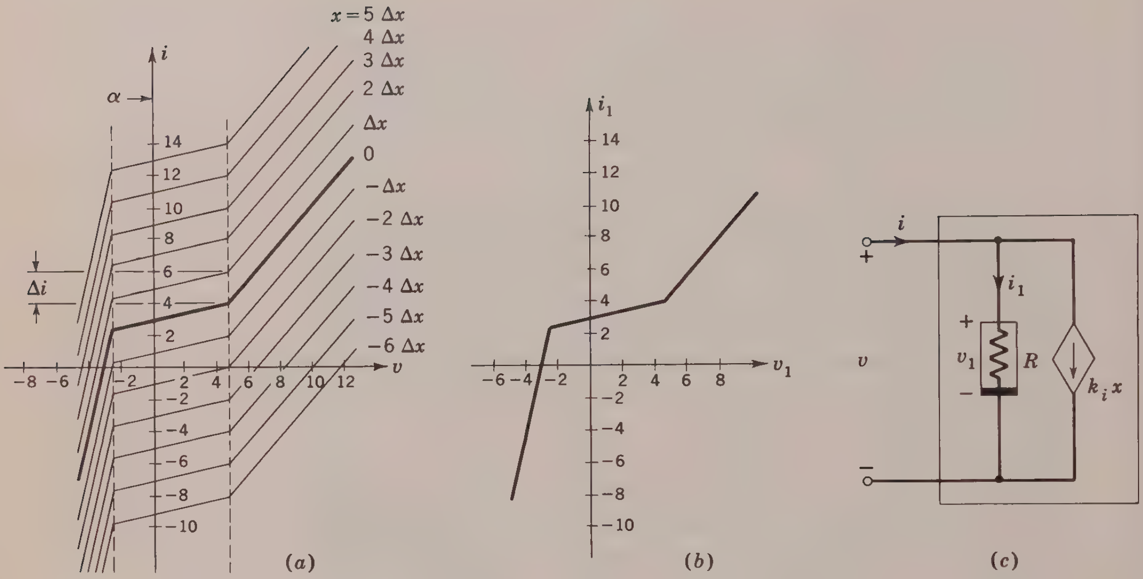
1. *When α coincides with the v axis.* An example of a system of v-i curves parallel with respect to the v axis is shown in Fig. 11-3*a*. The associated v-i-curve generator is shown in Fig. 11-3*b*. The model for this controlled resistor is shown in Fig. 11-3*c*, where the nonlinear resistor R is characterized by the v_1-i_1 curve in Fig. 11-3*b*. The controlling coefficient k_v of the controlled source is given by

$$k_v = \frac{\Delta v}{\Delta x} \tag{11-3}$$



(*a*) (*b*) (*c*)

Fig. 11-3. A family of v-i curves parallel with respect to the v axis and its model.



(a) (b) (c)

Fig. 11-4. A family of v-i curves parallel with respect to the i axis and its model.

2. *When α coincides with the i axis.* An example of a system of v-i curves parallel with respect to the i axis is shown in Fig. 11-4*a*. The associated v-i-curve generator is shown in Fig. 11-4*b*. The model for this controlled resistor is shown in Fig. 11-4*c*, where the nonlinear resistor R is characterized by the v-i-curve generator of Fig. 11-4*b*. The controlling coefficient k_i of the controlled source is given by

$$k_i = \frac{\Delta i}{\Delta x} \tag{11-4}$$

3. *When α is an arbitrary straight line.* Consider next the case where the straight line α makes an arbitrary angle φ with the vertical axis, as shown in Fig. 11-5*a*. Observe that if we rotate the v-i curves by $\varphi°$ in the counterclockwise direction or by $(90 - \varphi)°$ in the clockwise direction, the straight line α becomes the i axis or the v axis, respectively. Therefore, the model for this controlled resistor can be realized by first synthesizing a model N of the rotated v-i curves. The complete model is obtained by cascading N with a rotator with $\theta = -\varphi°$ or $\theta = (90 - \varphi)°$, respectively, as shown in Fig. 11-5*b*.

When a family of v-i curves does not satisfy properties 1 and 2, the above models are no longer valid. However, in most practical cases, it is easy to modify these models so that they are valid at least over some region of the v-i plane of interest. This modification can usually be achieved by adding appropriate nonlinear resistors. The principles behind this are similar to those presented in Chap. 8.

Exercise 1: Show that if the spacing between curves is not uniform, the models shown in Figs. 11-3, 11-4, and 11-5 are still valid provided the controlled source in each model is allowed to be nonlinear. Construct a simple example to illustrate the procedure.

Exercise 2: The rotator in the model of Fig. 11-5*b* is used simply for convenience. Show that after the model is obtained, the T-network or the π-network realization of the rotator can be *absorbed* with the network, thereby eliminating the rotator.

Exercise 3: Show that the family of v-i curves in Fig. 11-5*a* can be transformed into the form shown in Fig. 11-3*a* or Fig. 11-4*a* with the help of a negative resistor. From this, give a model analogous to that shown in Fig. 11-5*b*.

Exercise 4: Show that any *nonlinear* voltage-controlled voltage source described by $v_2 = f(v_1)$ can be realized by two *linear* controlled sources of the appropriate type and one *nonlinear* resistor.

Exercise 5: Repeat Exercise 4 for the remaining three types of nonlinear controlled sources.

11-4 SYNTHESIS OF GLOBAL MODELS FOR NONLINEAR THREE-TERMINAL RESISTORS

A three-terminal resistor is characterized by two sets of v-i curves. Although there are six distinct ways to represent these characteristics, they are equivalent in the sense that it is possible to transform one representation into another representation (see Sec. 3-2).

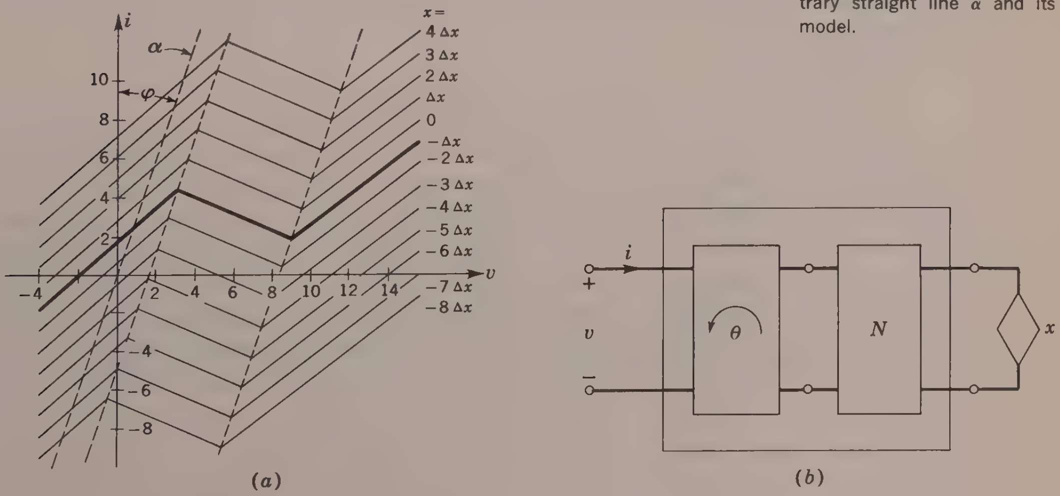


Fig. 11-5. A family of v-i curves parallel with respect to an arbitrary straight line α and its model.

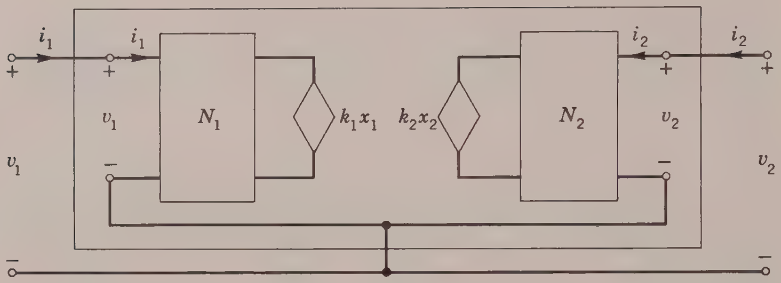


Fig. 11-6. Any three-terminal resistor can be modeled by two controlled resistors.

There is no loss of generality, therefore, to assume that the input characteristics are given in the v_1-i_1 plane with either i_2 or v_2 as the parameter. Similarly, there is no loss of generality to assume that the output characteristics are given in the v_2-i_2 plane with either i_1 or v_1 as the parameter. We shall for convenience denote the parameter i_2 or v_2 of the input characteristics by x_1. Similarly, we shall denote the parameter i_1 or v_1 of the output characteristics by x_2.

The simplest way to model a three-terminal resistor is to synthesize the input characteristics and the output characteristics independently by the controlled resistors N_1 and N_2 as shown in Fig. 11-6. After N_1 and N_2 are synthesized, we can usually apply the concepts of equivalent networks to further simplify the resulting model. Since the model shown in Fig. 11-6 is completely general, the problem of modeling a three-terminal resistor reduces to that of modeling controlled resistors. Hence, the principles presented in the preceding section are applicable to the modeling of three-terminal resistors as well. Let us consider two special but quite common cases.

11-4-1 A BASIC T-NETWORK MODEL

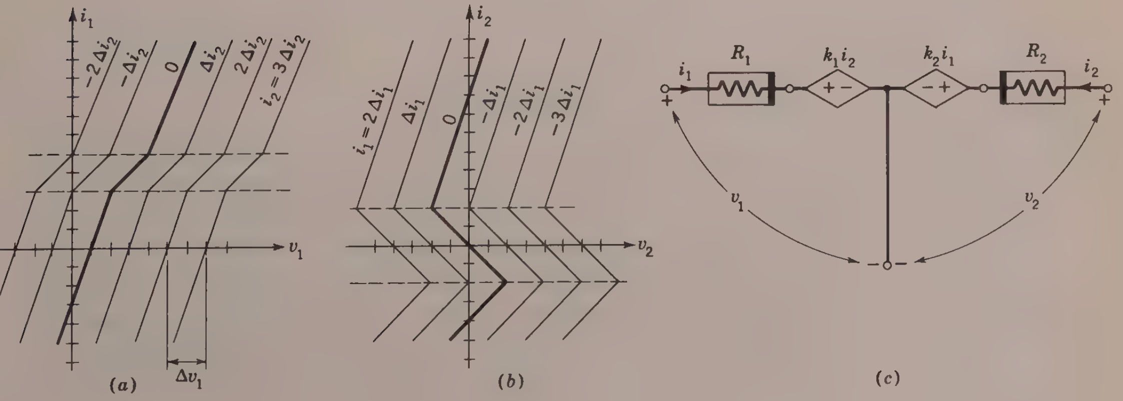
Consider the two sets of hypothetical v-i curves shown in Fig. 11-7*a* and 11-7*b*. Since these curves are parallel with respect to the voltage axis, they can be modeled by the T network shown in Fig. 11-7*c*. The resistor R_1 is characterized by the v_1-i_1-curve generator (corresponding to $i_2 = 0$) of the input characteristics. Similarly, the resistor R_2 is characterized by the v_2-i_2-curve generator (corresponding to $i_1 = 0$) of the output characteristics. The controlling coefficients k_1 and k_2 are given, respectively, by

$$k_1 = \frac{\Delta v_1}{\Delta i_2} \tag{11-5}$$

$$k_2 = \frac{\Delta v_2}{\Delta i_1} \tag{11-6}$$

Although the model shown in Fig. 11-7*c* is simple enough, the presence of the two current-controlled voltage sources (connected to a common node) suggests that we might be able to apply the *v*-shift theorem and eliminate one of the two controlled sources. However, in order to apply this theorem, the terminal voltage across the two controlled sources must be identical. We must, therefore, perform some transformation to make this possible. First, let us redraw the model of Fig. 11-7*c* into the equivalent form shown in Fig. 11-8*a*. These two circuits are equivalent because we have added and subtracted identical controlled sources to and from each of the two controlled sources of Fig. 11-7*c*. The reason for doing this is obvious once we realize that the circuit of Fig. 11-8*a* can be simplified to that shown in Fig. 11-8*b*. Observe that the controlled source $-k_1 i_1$ of Fig. 11-8*a* can be replaced by an equivalent linear negative resistance $-k_1$ (ohms) because the controlling variable i_1 is also the current flowing through the controlled source. Similarly, the controlled source $-k_1 i_2$ of Fig. 11-8*a* can be replaced by an equivalent negative resistance $-k_1$ (ohms) as shown in Fig. 11-8*b*.

We can now apply the *v*-shift theorem to shift the two identical controlled sources $k_1(i_1 + i_2)$ of Fig. 11-8*b* into the vertical branch shown in Fig. 11-8*c*. Observe that the controlling variable $(i_1 + i_2)$ of this controlled source is also the current entering its terminals. Therefore, once again we can replace this controlled source by a linear resistance k_1 (ohms) as shown in Fig. 11-8*d*.



(*a*) (*b*) (*c*)

Fig. 11-7. A set of three-terminal resistor characteristics and its representation by a T-network model containing only two-terminal resistors and controlled sources.

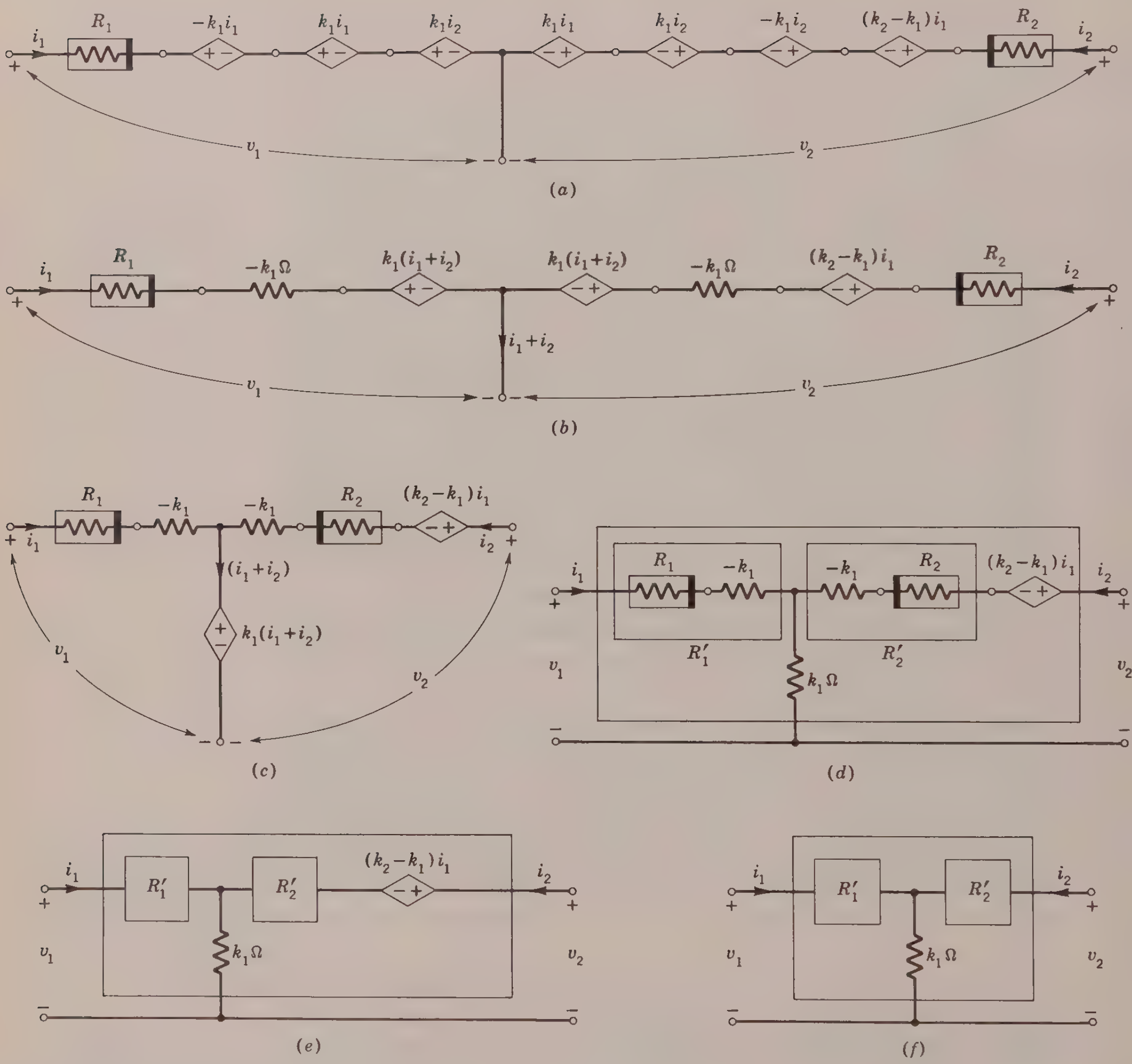


Fig. 11-8. The step-by-step procedure for arriving at an equivalent T-network model containing, at most, one controlled source.

We can simplify the model shown in Fig. 11-8*d* further by replacing the nonlinear resistor R_1 and the negative resistance $-k_1$ with a new nonlinear resistor R_1'. Similarly, we can replace the nonlinear resistor R_2 and the negative resistance $-k_1$ with a new nonlinear resistor R_2'. The resulting T-network model is shown in Fig. 11-8*e*. It is interesting to observe that in the special case where $k_1 = k_2$, this model reduces to a T network without the controlled source, as shown in Fig. 11-8*f*.

11-4-2 A BASIC π-NETWORK MODEL

Consider next the hypothetical v-i curves shown in Fig. 11-9*a* and *b*. Since these curves are parallel with respect to the current axis, they can be modeled by the π network shown in Fig. 11-9*c*. The resistor R_1 is characterized by the v_1-i_1-curve generator (corresponding to $v_2 = 0$) of the input characteristics. Similarly, the resistor R_2 is characterized by the v_2-i_2-curve generator (corresponding to $v_1 = 0$) of the output characteristics. The controlling coefficients k_1 and k_2 are given, respectively, by

$$k_1 = \frac{\Delta i_1}{\Delta v_2} \tag{11-7}$$

$$k_2 = \frac{\Delta i_2}{\Delta v_1} \tag{11-8}$$

Although the above model is simple enough, the presence of the two voltage-controlled current sources (connected to a common node) suggests that we might be able to apply the i-shift theorem and eliminate one of the two controlled sources. Hence, by a procedure dual to that used in the preceding section, we redraw the model of Fig. 11-9*c* into the equivalent model shown in Fig. 11-10*a*. This model can be further simplified by observing that the controlled source $k_1 v_1$ can be replaced by an equivalent resistance of $1/k_1$ (ohms). This is because the controlling parameter v_1 is also the voltage across the controlled source. Similarly, we can replace the controlled source $k_1 v_2$ by an equivalent resistance $1/k_1$ (ohms). The resulting model is shown in Fig. 11-10*b*.

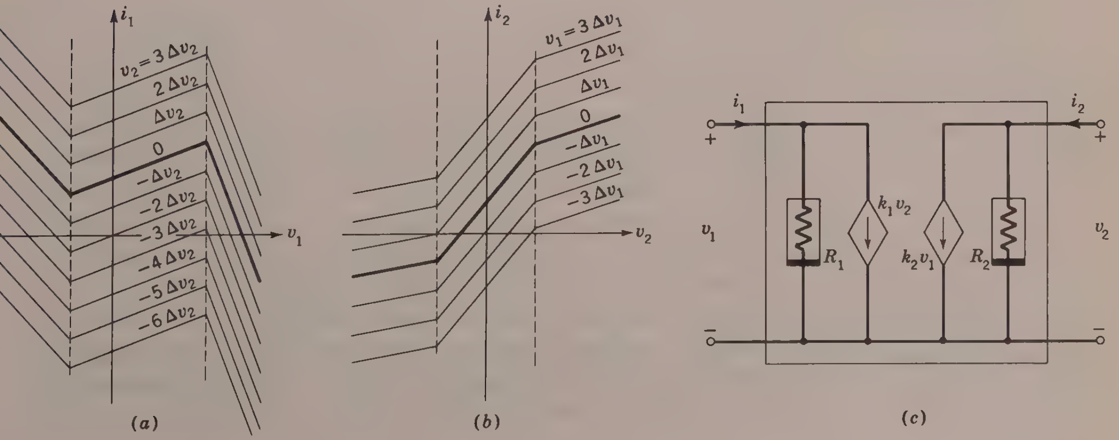


Fig. 11-9. A set of three-terminal resistor characteristics and its representation by a π-network model containing only two-terminal resistors and controlled sources.

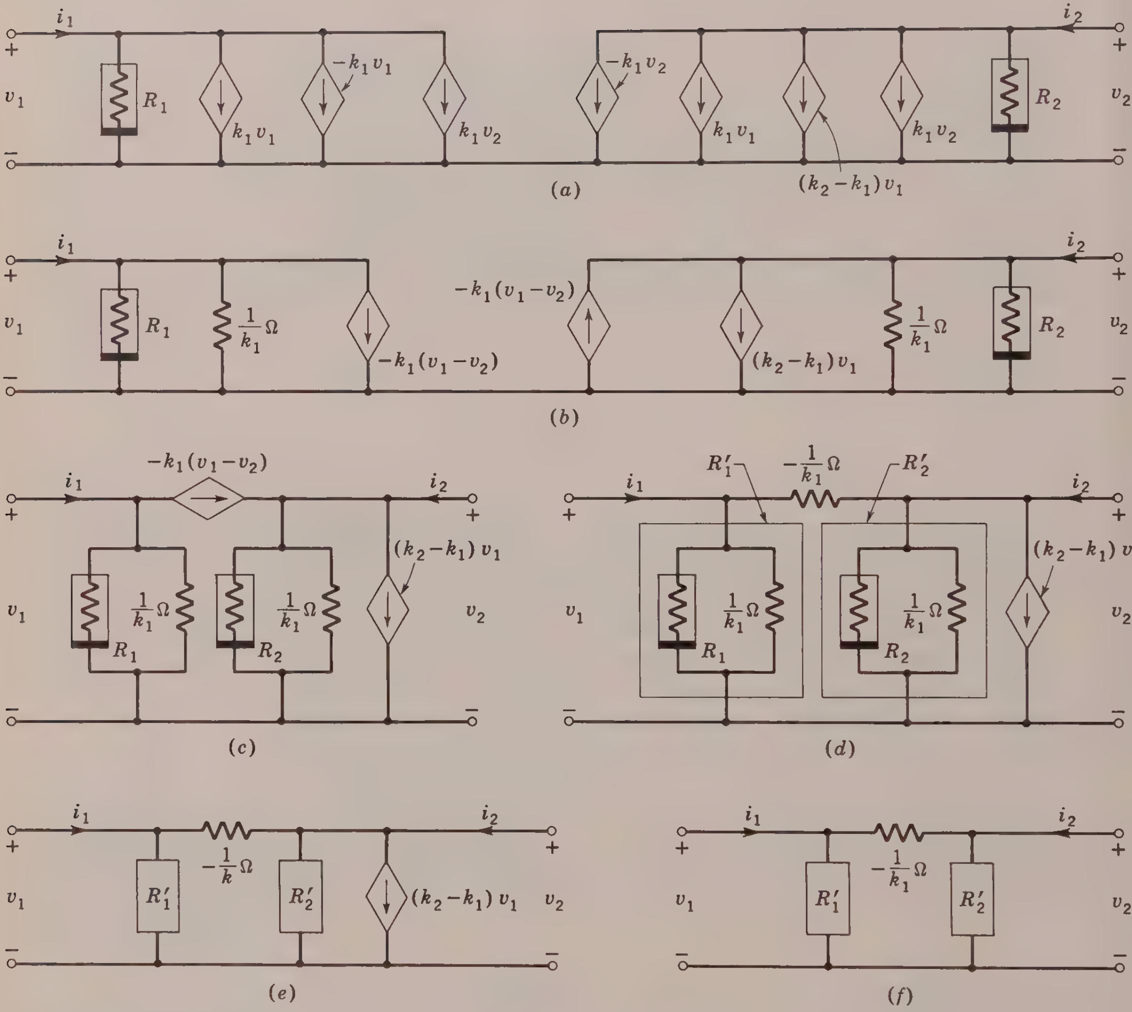


Fig. 11-10. The step-by-step procedure for arriving at an equivalent π-network model containing, at most, one controlled source.

We can now apply the i-shift theorem to shift the two controlled sources $-k_1(v_1 - v_2)$ into the horizontal branch shown in Fig. 11-10*c*. Observe that the controlling variable $(v_1 - v_2)$ of this controlled source is also the voltage across its terminals. Therefore, we can replace this controlled source by an equivalent negative resistance $-1/k_1$ (ohms) as shown in Fig. 11-10*d*.

We can simplify the model shown in Fig. 11-10*d* further by replacing the nonlinear resistor R_1 and the resistance $1/k_1$ with a new nonlinear resistor R_1'. Similarly, we can replace the nonlinear resistor R_2 and the resistance $1/k_1$ with a new nonlinear resistor

R_2'. The resulting π-network model is shown in Fig. 11-10*e*. It is interesting to observe that in the special case where $k_1 = k_2$, this model reduces to a π network without the controlled source, as shown in Fig. 11-10*f*.

The above T-network and π-network models are applicable to a large class of three-terminal resistors. In particular, any set of *v-i* curves which are parallel with respect to the voltage axis can be modeled by a T-network model.[1] Similarly, any set of *v-i* curves which are parallel with respect to the current axis can be modeled by a π-network model. In either case, one controlled source at most is required.

If the characteristics of a given three-terminal resistor are more complicated than those considered above, we can often construct a good model by merely adding appropriate nonlinear resistors and rotators to the two basic models obtained in this section. Some of the more common techniques for modifying these models are given in the next section.

Exercise 1: Give a geometrical interpretation of the effect of the condition $k_1 = k_2$ on the characteristic curves of the T-network model shown in Fig. 11-8*f*, and of the π-network model shown in Fig. 11-10*f*.

Exercise 2: Give a systematic procedure for finding the parameters and *v-i* curves of the T-network and the π-network models directly from the given set of characteristic curves.

11-5 SOME USEFUL TECHNIQUES FOR MODIFYING MODELS

There are many nonlinear resistors whose *v-i* curves appear only in a certain region of the *v-i* plane. For example, the *v-i* curves of any passive resistor can appear only in the first and third quadrants.[2] Since the *v-i* curves of the various models presented in the preceding sections cover the entire *v-i* plane, it is necessary to modify these models so that their *v-i* curves are constrained to appear only in some desired region of the *v-i* plane. We shall now present techniques to implement some of the more basic network constraints.

11-5-1 CONSTRAINING *v-i* CURVES TO HALF-PLANES

Any straight line in the *v-i* plane separates it into two regions, each of which is called a *half-plane*. Suppose it is required to constrain the *v-i* curves of a controlled-resistor model to appear only in a half-plane. Since we can always use the rotator to rotate the *v-i*

---

[1] If the system of curves is not uniformly spaced, the model is still valid, provided nonlinear controlled sources are allowed.

[2] Review Sec. 1-9 for the significance of passivity.

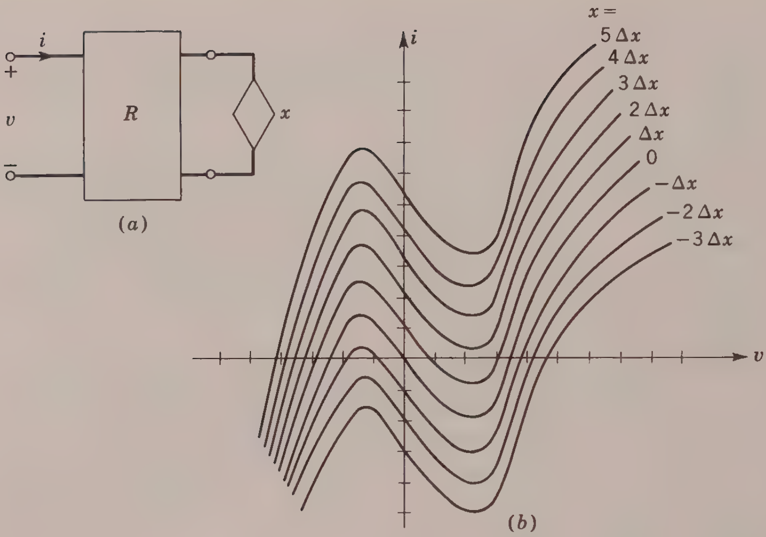


Fig. 11-11. The v-i curves of a hypothetical controlled-resistor model.

curves if necessary, there is no loss of generality to assume that the half-plane is defined either by a vertical line or by a horizontal line. Hence, we shall consider only the half-planes $v \geq E$, $v \leq E$, $i \geq I$, or $i \leq I$. To be specific, let us suppose the controlled-resistor model shown in Fig. 11-11*a* is characterized by the v-i curves shown in Fig. 11-11*b*.

A simple technique for constraining the v-i curves to the half-plane $v \geq E$ is shown in Fig. 11-12*a*. The resulting v'-i' curves shown in this figure can be readily verified graphically. We can also verify this analytically by observing that since the voltage across an ideal diode cannot be positive, $v_d = E - v' \leq 0$. But this implies that $v' \geq E$ and the diode can be replaced by an open circuit for all $v' > E$. Therefore, the v'-i' curves are identical with the v-i curves for $v' > E$. It is important to observe, however, that the resulting v'-i' curves contain the *lower* portion of the vertical line $v' = E$. Since this portion of the curves need not be present in the original v-i curves, the model shown in Fig. 11-12*a* is generally valid only in the region $v' > E$. Observe that when $E = 0$, a special but rather common case, the v-i curves are constrained to the right half-plane. In order to constrain the v-i curves to the half-plane $v \leq E$, all we need to do is to reverse the diode terminals as shown in Fig. 11-12*b*.

By dual procedure we can constrain the v-i curves to the half-plane $i \geq -I$ by the circuit shown in Fig. 11-12*c*. Similarly, by reversing the diode terminals as shown in Fig. 11-12*d*, we can constrain the v-i curves to the half-plane $i \leq -I$.

11-5-2 CONSTRAINING v-i CURVES TO QUADRANTS

We can easily constrain the v-i curves to appear in any quadrant by constraining the curves into two separate half-planes. For example, to constrain the v-i curves of Fig. 11-11*b* to the first quadrant, we first constrain the v-i curves to the right half-plane $v \geq 0$ by diode D_1 as shown in Fig. 11-13*a*. This is then followed by constraining the resulting v-i curves to the upper half-plane $i \geq 0$ by diode D_2. Alternately, we can first constrain the v-i curves to the upper half-plane $i \geq 0$, then follow by constraining the v-i curves to the right half-plane $v \geq 0$, as shown in Fig. 11-13*b*. Observe that the only difference between these two models is that the v'-i' curves of Fig. 11-13*a* contain a portion of the v' axis, whereas the v'-i' curves of Fig. 11-13*b* contain a portion of the i' axis.

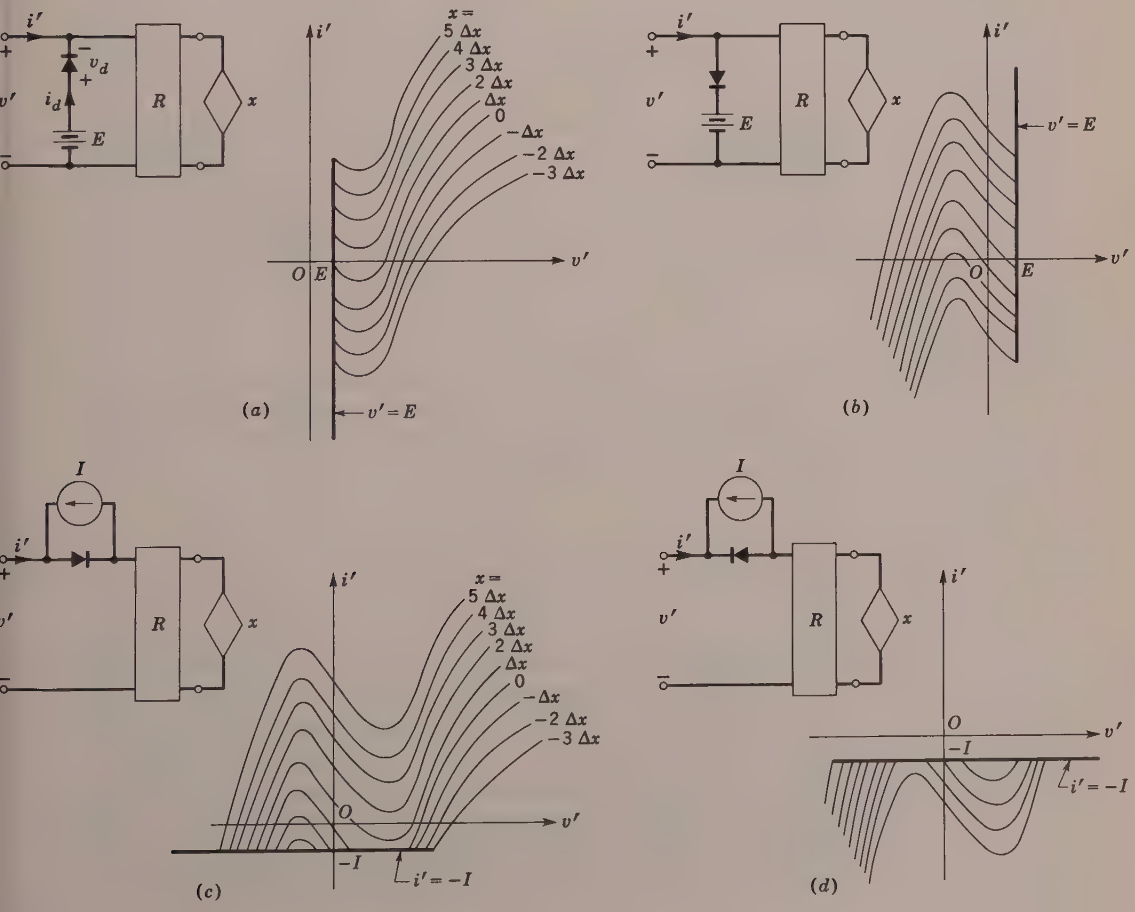
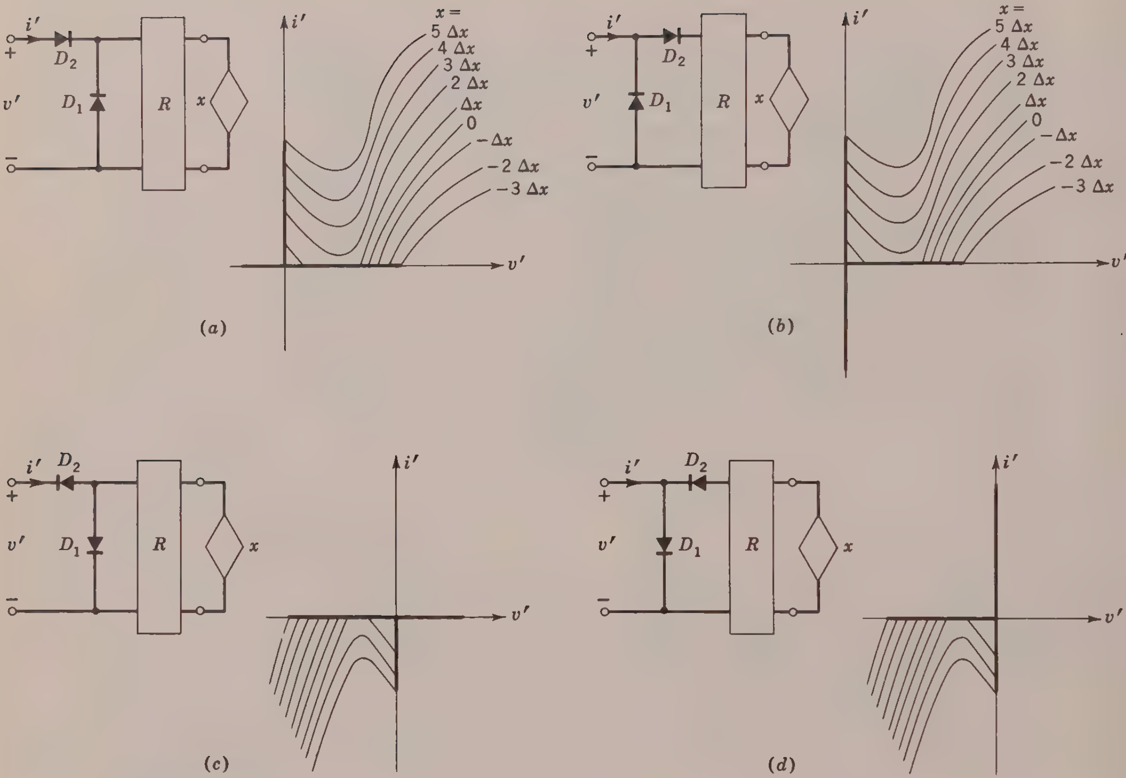


Fig. 11-12. Techniques for constraining v-i curves to half-planes.



(a)

(b)

(c)

(d)

Fig. 11-13. Techniques for constraining v-i curves to quadrants.

By similar procedures, we can constrain the v-i curves to appear only in the second, third, or fourth quadrant. For example, the circuits shown in Fig. 11-13*c* and *d* can be used to constrain the v-i curves to the third quadrant.

11-5-3 CONSTRAINING v-i CURVES TO THE FIRST AND THIRD QUADRANTS

Another common modeling problem is to constrain the v-i curves of a given model to appear only in the first and the third quadrants. If the v-i curves are symmetrical with respect to the origin, the simplest circuit for accomplishing this task is shown in Fig. 11-14*a*. To verify that the v-i curves are indeed constrained to these quadrants, as shown in Fig. 11-14*b*, let us apply the v-shift theorem to shift the input voltage source v' into the diode branches and observe that when $v' \geq 0$, diodes D_1 and D_2 are short-circuited, while diodes D_3 and D_4 are open-circuited. The opposite case occurs when $v' < 0$. In either case, it is impossible to obtain points in the second and the fourth quadrants. A careful examination of this circuit shows that the diode lattice network plays the same role as the switching circuit considered earlier in Fig. 10-11*a*. Hence, if the given v-i curves are not symmetrical with respect to the origin, they will be transformed into a set of symmetrical v'-i' curves, where the curves in the third quadrant are obtained by rotating the curves in the first quadrant by 180°. Another circuit must therefore be used in order to constrain a *nonsymmetrical* set of v-i curves to the first and third quadrants. One such circuit consists of a parallel (series) connection of two separate models. One model realizes the v-i curves in the first quadrant by the technique shown in Fig. 11-13*a* (Fig. 11-13*b*). The other model realizes the v-i curves in the third quadrant by the technique shown in Fig. 11-13*c* (Fig. 11-13*d*).

Exercise 1: Show a method for constraining the v-i curves to the right of the 45° line ($v = i$).

Exercise 2: By defining another two-terminal resistor analogous to an ideal diode, show that it is possible to constrain the v-i curves to the second or the fourth quadrant. Why is it all right for us to define any two-terminal resistor that we need for implementing a given constraint without questioning its realizability and physical interpretation?

Exercise 3: Show a method for constraining the v-i curves to the second *and* the fourth quadrants. HINT: See Exercise 2.

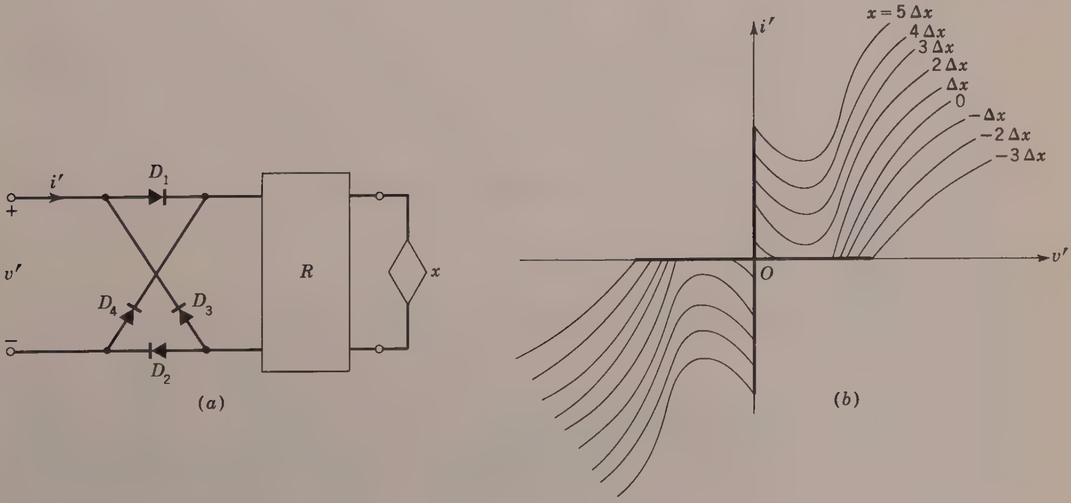


Fig. 11-14. Techniques for constraining symmetrical (with respect to the origin) v-i curves to the first and third quadrants.

11-6 MODELS OF PRACTICAL THREE-TERMINAL RESISTORS

Let us now apply the principles presented in the preceding sections to synthesize models for some of the more common three-terminal resistors. Since no new concept is involved, we shall be brief and leave some of the details to the problems. Since the models to be synthesized in this section will be referred to again many times in the future, it is convenient to identify each model by a name. Indeed, the present section can be considered as an appendix, listing the models of various common three-terminal resistors. To enhance its usefulness, we shall first synthesize an accurate global model for each three-terminal resistor considered in this section. For most practical purposes, these global models are actually more complicated than necessary. For this reason, we shall further reduce each global model into simpler models.

11-6-1 VACUUM TRIODE MODELS

The standard symbol normally associated with vacuum triodes is shown in Fig. 11-15*a*. The input and output characteristics of a typical vacuum triode can be represented to a fair degree of precision by the piecewise-linear characteristics shown in Fig. 11-15*b* and *c*, respectively. Since these characteristics appear only in the first quadrant, we shall first find two simple models whose *v-i* curves agree with the desired input and output characteristics in this quadrant. The *v-i* curves in the remaining quadrants are immaterial since we shall later forbid any curve from appearing in these regions. Clearly, the two simple linear models shown in Fig. 11-15*d* and *e* have a set of *v-i* curves which agree with the triode characteristics in the first quadrant. The controlling coefficients γ and μ can be computed from the given characteristics by[1]

$$\gamma = \left|\frac{\Delta V_G}{\Delta V_P}\right| \qquad \text{(from input characteristics)} \tag{11-9}$$

and

$$\mu = \left|\frac{\Delta V_P}{\Delta V_G}\right| \qquad \text{(from output characteristics)} \tag{11-10}$$

The next step is to constrain the *v-i* curves so that they appear only in the first quadrant. Since the triode *v-i* curves contain a portion of the horizontal axis, we choose the technique of Fig. 11-13*a* to modify the models as shown in Fig. 11-15*f* and *g*. Com-

---

[1] The controlling coefficient μ is usually called the *amplification factor.* For global models the amplification factor may be assumed to be a constant, even though it actually varies slightly from point to point. This is because when we are dealing with large signals, as in global models, the small variation in μ is negligible.

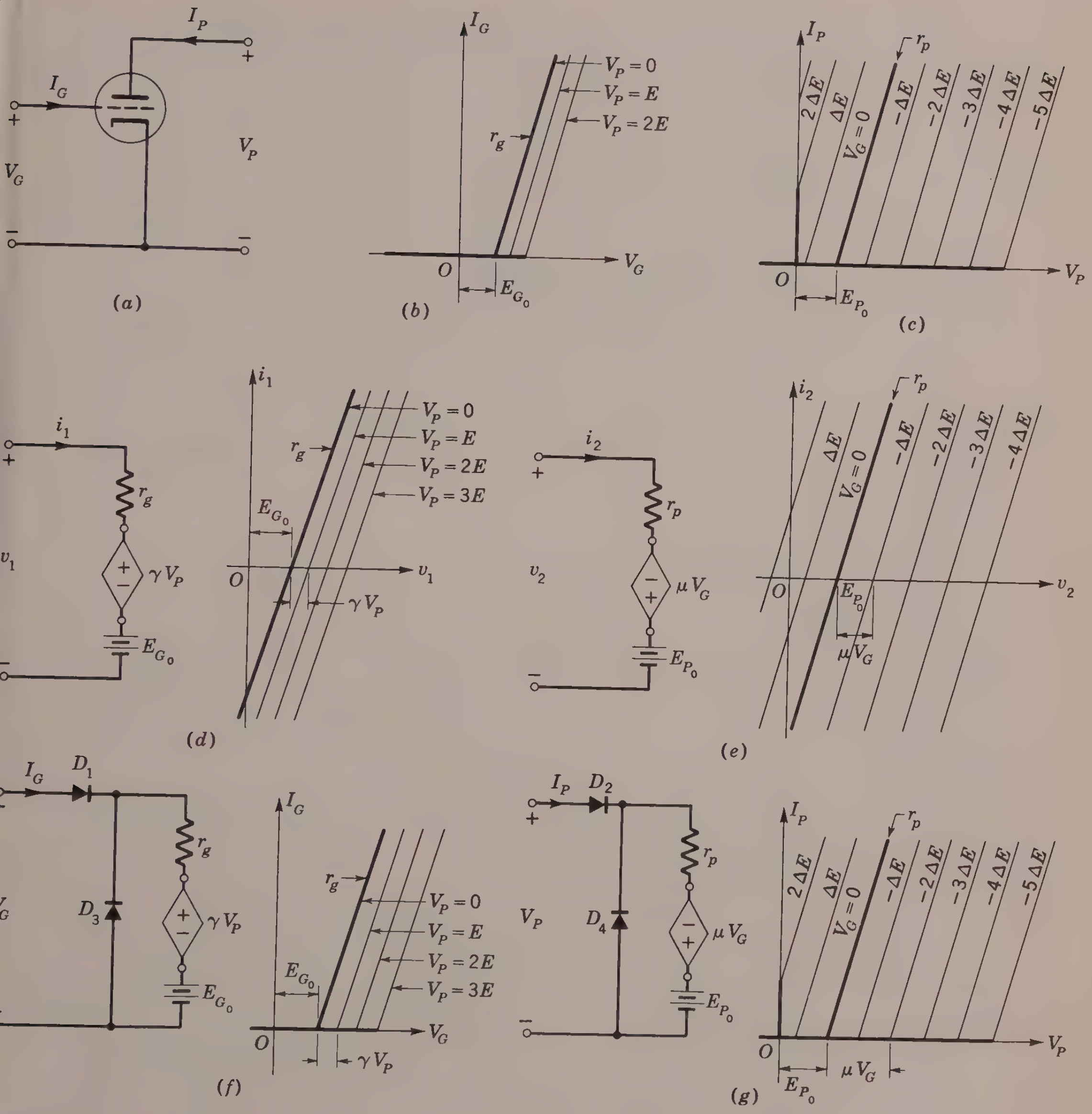
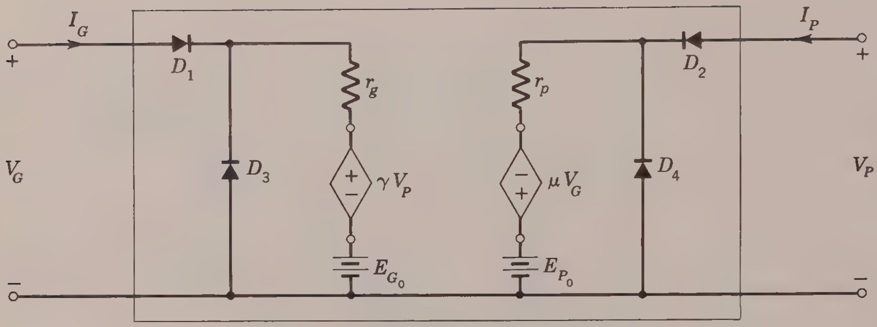


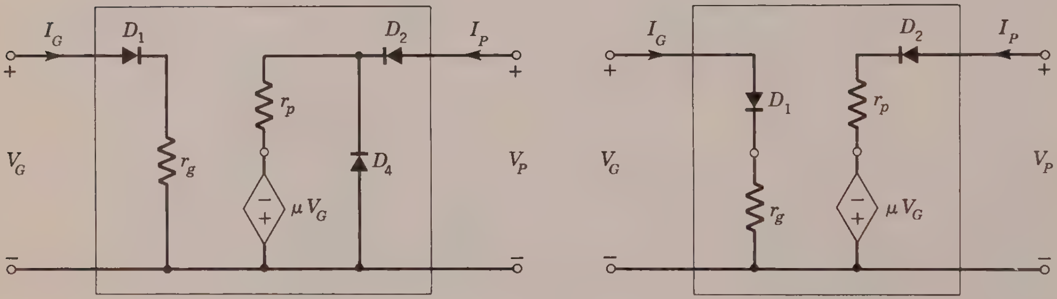
Fig. 11-15. The step-by-step procedure for synthesizing a global vacuum triode model.

bining these two controlled-resistor models, we obtain our first triode model A as shown in Fig. 11-16*a*.

The typical values for γ, E_{G_0}, and E_{P_0} are generally quite small and may be neglected for most practical purposes. Hence, setting the coefficients γ, E_{G_0}, and E_{P_0} in triode model A to zero, we obtain the simpler triode model B shown in Fig. 11-16*b*. Ob-



(*a*) Triode model *A*



(*b*) Triode model *B*

(*c*) Triode model *C*

Fig. 11-16. Three models of a vacuum triode.

serve that the ideal diode D_3 of triode model A has also been removed since it is no longer needed in the absence of the controlled source γV_P.

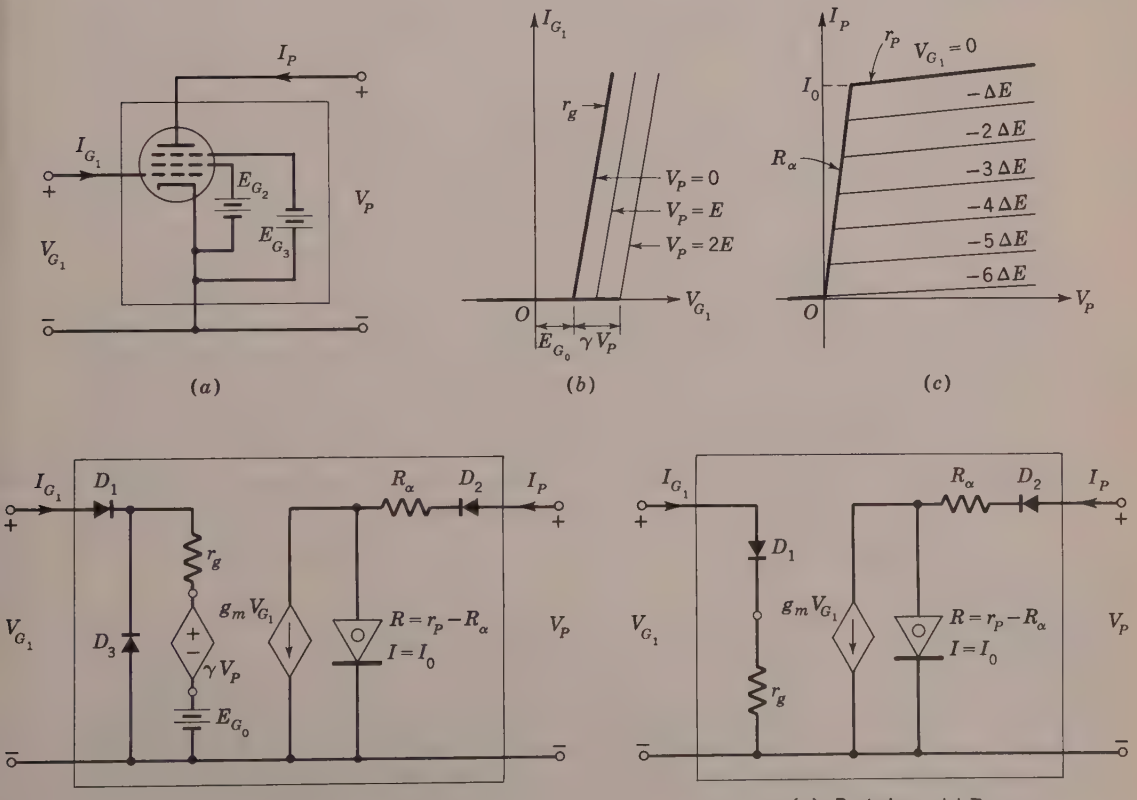
The triode model B is still a global model. In practice, the vacuum triode is usually operated within the first quadrant. Under this condition, we can eliminate the ideal diode D_4, whose only purpose is to prevent the output characteristics from appearing in the second quadrant, and obtain the triode model C shown in Fig. 11-16*c*. Observe that this model is a "local model" because for positive values of V_G, the output curves of this model will contain curves in the second quadrant, and they do not represent the actual triode output characteristics. Hence, triode model C is valid only if the operating points of the network lie in the first quadrant.

11-6-2 VACUUM PENTODE MODELS

The standard symbol normally associated with vacuum pentodes is shown in Fig. 11-17*a*. Although the pentode is a five-terminal resistor, it is usually operated as a three-terminal resistor. The input and output characteristics of a typical vacuum pentode (with E_{G_2} and E_{G_3} fixed by batteries) can be represented to a fair degree of precision by the piecewise-linear characteristics shown in Fig. 11-17*b* and *c*, respectively. Since the pentode input characteristics are identical with the triode input characteristics, the same controlled-resistor model shown in Fig. 11-15*f* is applicable here.

Since the pentode output characteristics are uniformly spaced, and since all breakpoints lie in a straight line corresponding to the resistance R_α, we can apply the principles presented in Sec. 11-3 to synthesize an output-characteristic model. Combining this with the input-characteristic model, we obtain the pentode model A shown in Fig. 11-17*d*. For most practical purposes, the values of γ and E_{G_0} are small and may be neglected, in which case we obtain the simpler pentode model B as shown in Fig. 11-17*e*.

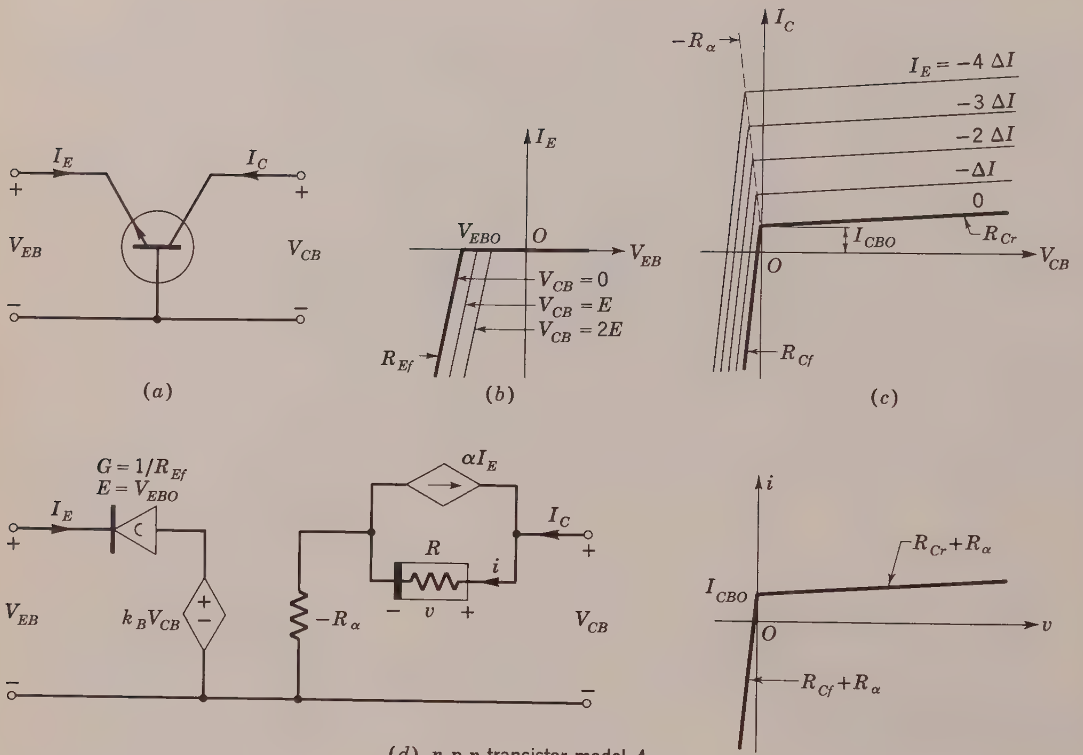
Fig. 11-17. Two models of a vacuum pentode.



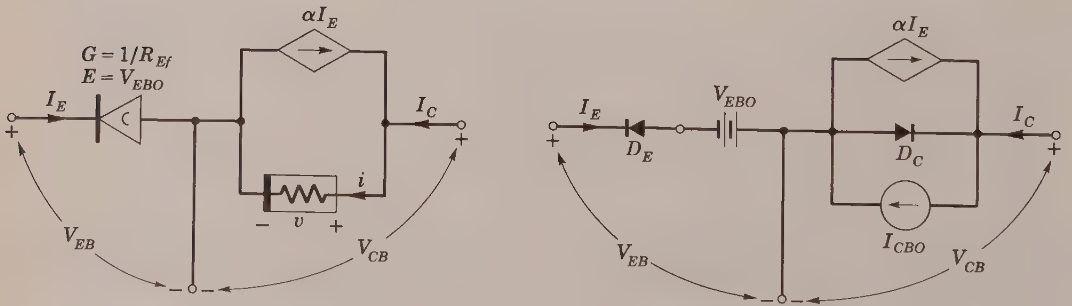
(*a*) (*b*) (*c*)

(*d*) Pentode model *A*

(*e*) Pentode model *B*



(d) n-p-n transistor model A_B



(e) n-p-n transistor model B_B

(f) n-p-n transistor model C_B

Fig. 11-18. Three models of an n-p-n transistor in the common-base configuration.

11-6-3 n-p-n TRANSISTOR MODELS

The standard symbol normally associated with an n-p-n transistor in the common-base configuration is shown in Fig. 11-18a. Its input and output characteristics can be represented quite accurately by the piecewise-linear v-i curves shown in Fig. 11-18b and c, respectively. Applying the principles in Sec. 11-3, we obtain the global *n-p-n* transistor model $\mathrm{A_B}$ as shown in Fig. 11-18*d*.[1] The controlling coefficients k_B and α are given by

$$k_B = \left|\frac{\Delta V_{EB}}{\Delta V_{CB}}\right| \quad \text{(from third quadrant of the input characteristics)} \tag{11-11}$$

$$\alpha = \left|\frac{\Delta I_C}{\Delta I_E}\right| \quad \text{(from first quadrant of the output characteristics)} \tag{11-12}$$

For most practical purposes the values of k_B and R_α are negligible, in which case we obtain the simpler but still very realistic *n-p-n* transistor model $\mathrm{B_B}$ as shown in Fig. 11-18*e*.

In many large-signal transistor circuits such as *multivibrators*, we can further simplify the above model without introducing significant errors. The simplification consists of assuming $R_{Ef} \approx 0$, $R_{Cf} \approx 0$, and $R_{Cr} \approx \infty$. Under these assumptions the concave resistor reduces to an ideal diode in series with a battery, and the nonlinear resistor R reduces to an ideal diode in parallel with a current source I_{CBO}. The resulting *n-p-n* transistor model $\mathrm{C_B}$ is shown in Fig. 11-18*f*.

The parameters associated with models $\mathrm{A_B}$, $\mathrm{B_B}$, and $\mathrm{C_B}$ must be obtained from the transistor characteristics in the common-base configuration. However, the majority of transistor manufacturers furnish their transistor characteristics in the common-emitter configuration because these characteristics are easier to obtain in practice. While it is possible to find the parameters of models $\mathrm{A_B}$, $\mathrm{B_B}$, and $\mathrm{C_B}$ from the common-emitter characteristics directly, the procedure is tedious, and the result is often not very accurate. It is more convenient, therefore, to synthesize separate transistor models for the common-emitter configuration.

The symbol and characteristics of a typical *n-p-n* transistor in the common-emitter configuration are shown in Fig. 11-19*a* to *c*. Applying the principles in Sec. 11-3, we obtain the *n-p-n* transistor model $\mathrm{A_E}$ shown in Fig. 11-19*d*.[2] The controlling coefficients k_E, k_C, and β are given by

$$k_E = \left|\frac{\Delta V_{BE}}{\Delta V_{CE}}\right| \quad \text{(from first quadrant of the input characteristics)} \tag{11-13}$$

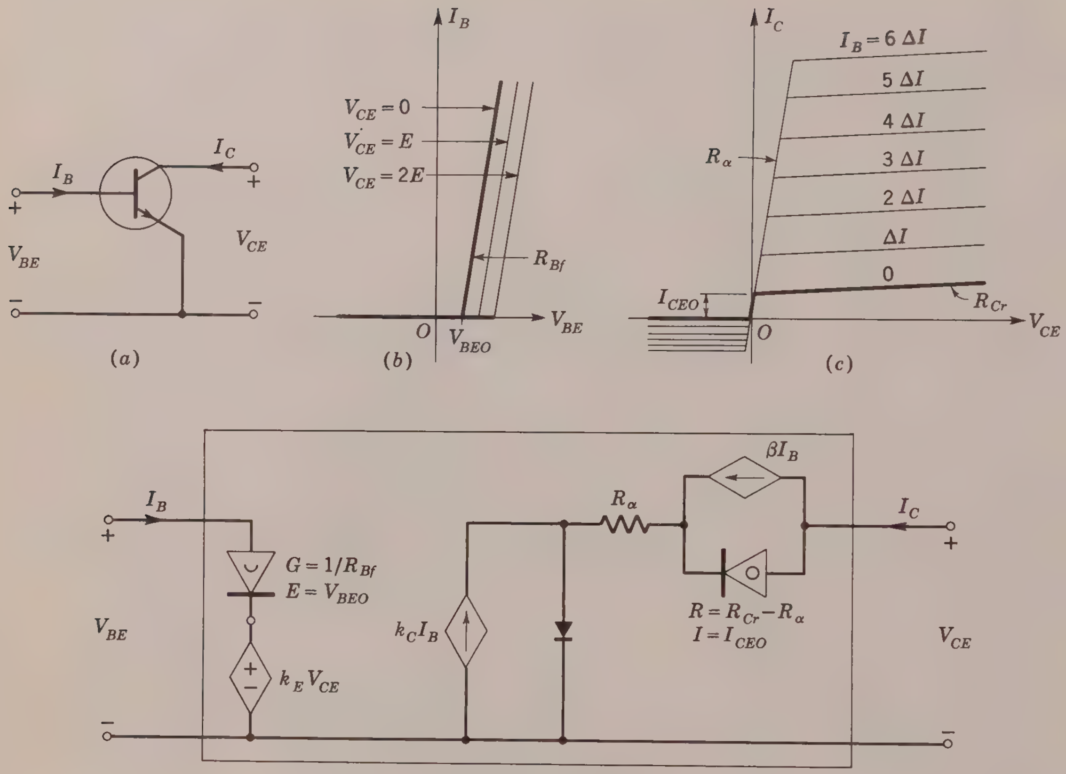
$$K_C = \left|\frac{\Delta I_C}{\Delta I_B}\right| \quad \text{(from third quadrant of the output characteristics)} \tag{11-14}$$

$$\beta = \left|\frac{\Delta I_C}{\Delta I_B}\right| \quad \text{(from first quadrant of the output characteristics)} \tag{11-15}$$

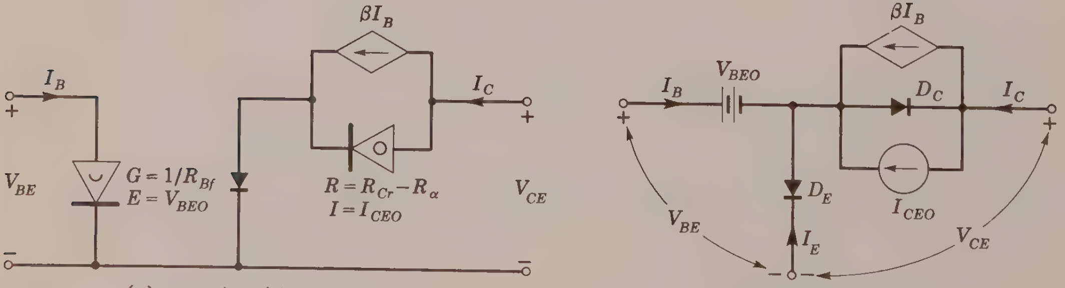
---

[1] The subscript B implies that the characteristic curves are obtained with the transistor in the *common-base* configuration.

[2] The subscript E denotes the common-emitter configuration.



(*d*) *n-p-n* transistor model A_E



(*e*) *n-p-n* transistor model B_E

(*f*) *n-p-n* transistor model C_E

Fig. 11-19. Three models of an *n-p-n* transistor in the common-emitter configuration.

For most practical purposes, the values of k_E, k_C, and R_α are negligible, in which case we obtain the simpler but still very realistic *n-p-n* transistor model B_E shown in Fig. 11-19*e*.

In many large-signal transistor circuits we can further modify model B_E to obtain the still simpler *n-p-n* transistor model C_E shown in Fig. 11-19*f*.[1]

[1] Observe that model C_E is not a special case of model B_E. It is, however, approximately equivalent to setting $R_{Bf} \approx 0$ and $R_{Cr} \approx \infty$ in model B_E.

It is important to observe that because different approximations are involved, models A_B, B_B, and C_B are only *approximately* equivalent to models A_E, B_E, and C_E. In many practical problems it is useful to find a relationship between α and β and a relationship between I_{CBO} and I_{CEO}. Since these parameters pertain only to the collector *v-i* curves in the first quadrant, we may replace the diode D_C in models C_E and C_B by open circuits. This allows us to write the following two KCL equations from models C_B and C_E, respectively:

$$\begin{aligned} I_C &= I_{CBO} - \alpha I_E \\ &= I_{CBO} - \alpha[-(I_B + I_C)] \qquad (11\text{-}16) \\ I_C &= I_{CEO} + \beta I_B \qquad (11\text{-}17) \end{aligned}$$

Solving for I_C from Eq. (11-16) and equating it to Eq. (11-17), we obtain

$$I_{CEO} + \beta I_B = \frac{I_{CBO}}{(1-\alpha)} + \frac{\alpha}{(1-\alpha)} I_B \qquad (11\text{-}18)$$

Equating the corresponding terms, we obtain

$$\beta = \frac{\alpha}{1-\alpha} \qquad (11\text{-}19)$$

$$I_{CEO} = \frac{I_{CBO}}{(1-\alpha)} \qquad (11\text{-}20)$$

Conversely, we can solve for α and β from Eqs. (11-19) and (11-20) to obtain

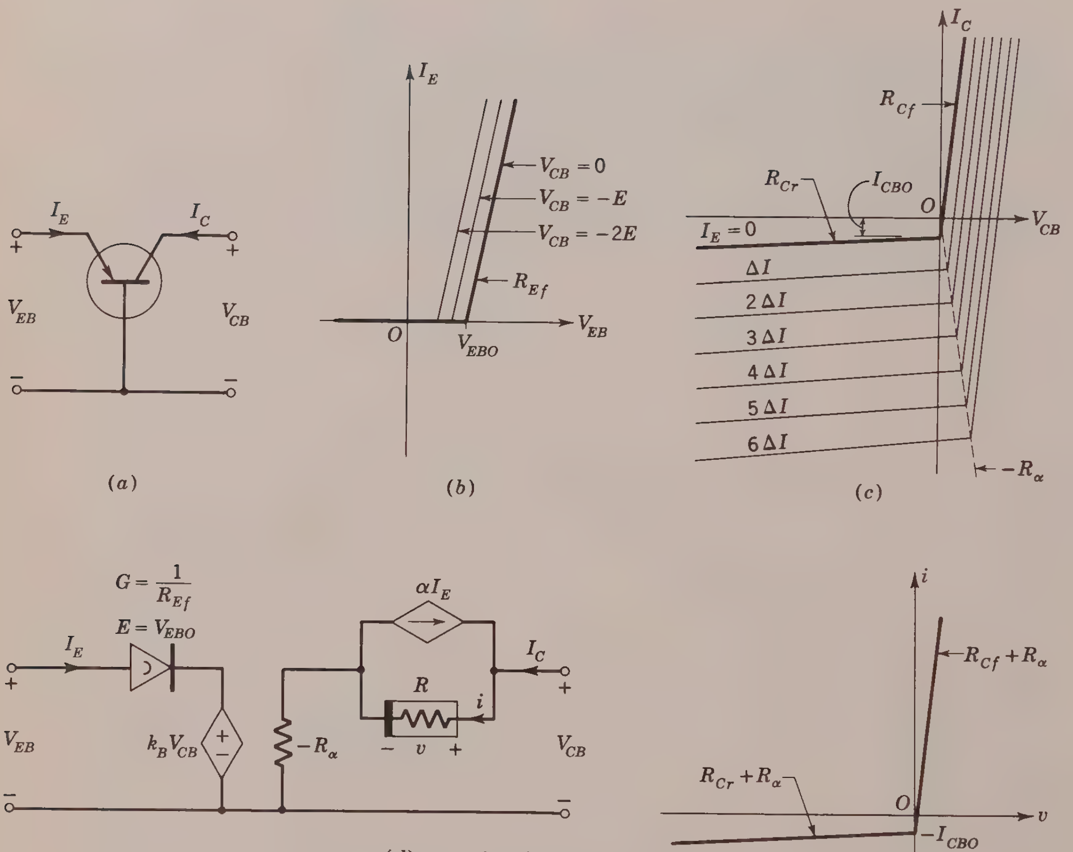
$$\alpha = \frac{\beta}{(\beta+1)} \qquad (11\text{-}21)$$

$$I_{CBO} = \frac{I_{CEO}}{(\beta+1)} \qquad (11\text{-}22)$$

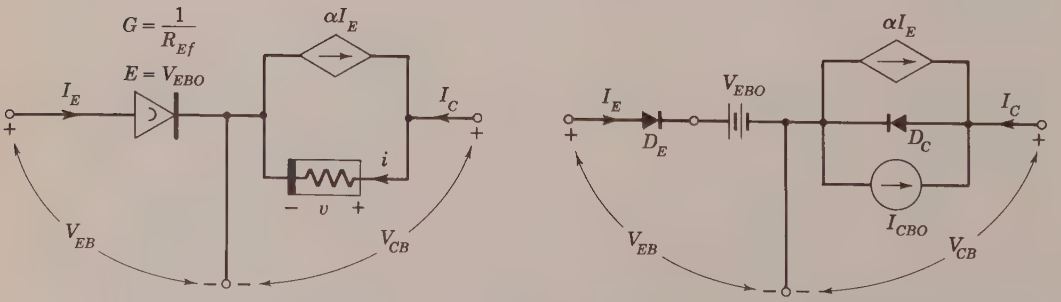
Since $\beta \gg 1$ for most transistors, Eq. (11-22) can be written as

$$I_{CEO} \approx \beta I_{CBO} \qquad (11\text{-}23)$$

This equation shows that the current source I_{CEO} in the common-emitter model plays a much more significant role than the current source I_{CBO} in the common-base model. Because of this, we may generally neglect I_{CBO} but not I_{CEO}.



(d) p-n-p transistor model A_B



(e) p-n-p transistor model B_B

(f) p-n-p transistor model C_B

Fig. 11-20. Three models of a p-n-p transistor in the common-base configuration.

11-6-4 *p-n-p* TRANSISTOR MODELS

By similar procedures, we can construct models for *p-n-p* transistors whose symbol is shown in Fig. 11-20*a*. However, since we already have the *n-p-n* transistor models, we can more easily obtain the *p-n-p* transistor models by applying the complementary-network theorem. This consists of interchanging the terminals of all nonlinear resistors and *independent* sources of the *n-p-n* transistor models. Hence, for the common-base *p-n-p* transistor characteristics shown in Fig. 11-20*b* and *c*, we immediately obtain the corresponding *p-n-p* transistor models A_B, B_B, and C_B as shown in Fig. 11-20*d* to *f*. Similarly, for the common-emitter *p-n-p* transistor in Fig. 11-21*a* and its characteristics in Fig. 11-21*b* and *c*, we obtain the corresponding *p-n-p* transistor models A_E, B_E, and C_E as shown in Fig. 11-21*d*, *e*, and *f*, respectively. Clearly, the controlling coefficients for the various controlled sources are given by the same expressions defined for *n-p-n* transistors. Similarly, the relationships between α, β, I_{CBO}, and I_{CEO} are also given by the same expressions derived earlier for *n-p-n* transistors.

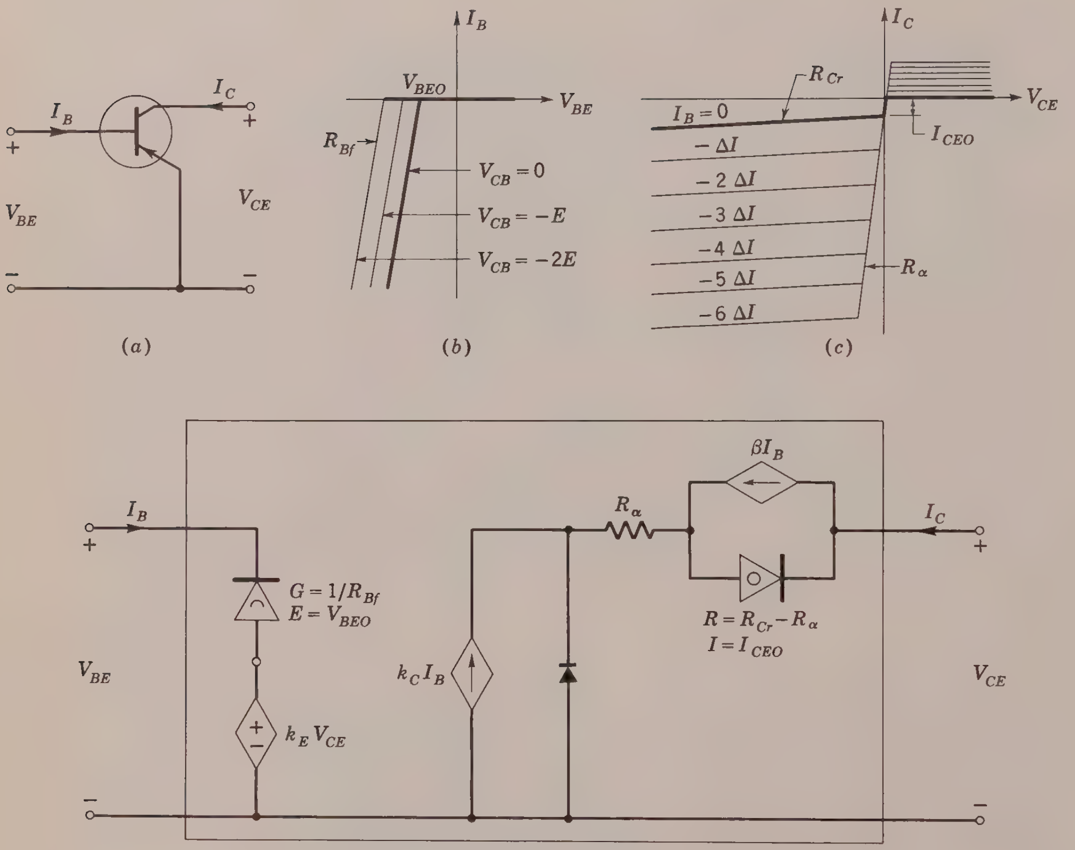
11-6-5 *n*-CHANNEL FET MODELS

The symbol and piecewise-linearized characteristics of a typical *n*-channel FET are shown in Fig. 11-22*a* to *c*. Since the input characteristics are independent of the output variables I_D and V_{DS}, it can be modeled by a two-terminal resistor R_G with the *v-i* curve shown in Fig. 11-22*b*. The output characteristics can be modeled by one controlled source and two nonlinear resistors. The complete *n*-channel FET model A is shown in Fig. 11-22*d*. The v_1-i_1 curve of R_1 is shown in Fig. 11-22*e*, and the v_2-i_2 curve of R_2 is shown in Fig. 11-22*f*. The controlling coefficient k is given by

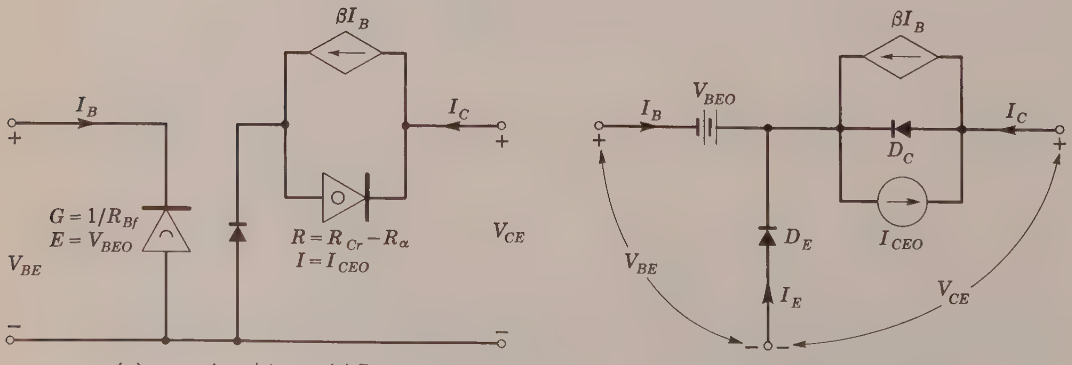
$$k = \frac{\Delta I_D}{\Delta V_{GS}} \tag{11-24}$$

11-6-6 *p*-CHANNEL FET MODELS

The symbol of a *p*-channel FET is shown in Fig. 11-23*a*. The characteristics of a *p*-channel FET differ from the characteristics of an *n*-channel FET only by a negative sign in the variables I_G, V_{GS}, I_D, and V_{DS}. Hence, applying the complementary-network theorem once again, we obtain the *p*-channel FET model A shown in Fig. 11-23*b*.



(d) *p-n-p* transistor model A_E



(e) *p-n-p* transistor model B_E

(f) *p-n-p* transistor model C_E

Fig. 11-21. Three models of a *p-n-p* transistor in the common-emitter configuration.

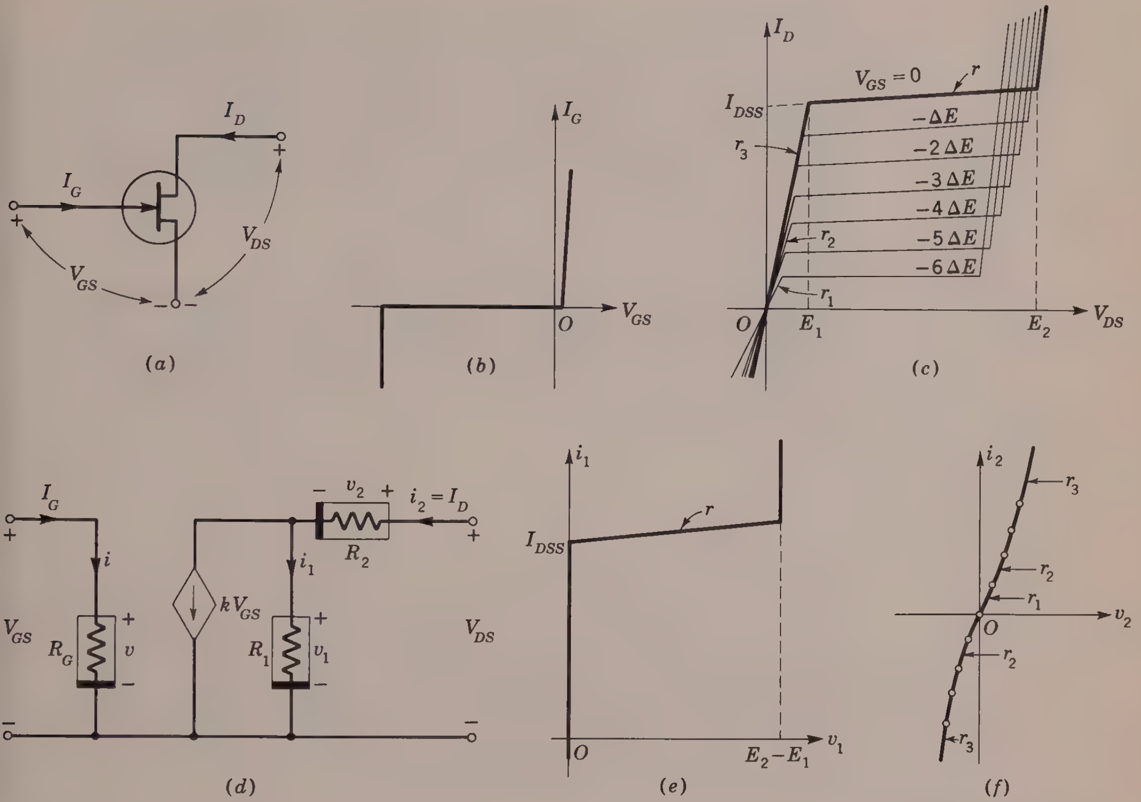


Fig. 11-22. An *n*-channel FET model A.

Exercise 1: (*a*) Find all pertinent parameters associated with the triode model A for triode type 2NXX22. (*b*) Sketch the characteristic curves predicted by this model, as well as the triode models B and C as extracted from model A. (*c*) Compare the accuracy of each model.

Exercise 2: (*a*) Find all pertinent parameters associated with the pentode model A for pentode type 2NXX24. (*b*) Sketch the characteristic curves predicted by this model and its simpler version given by pentode model B. (*c*) Compare the accuracy of each model.

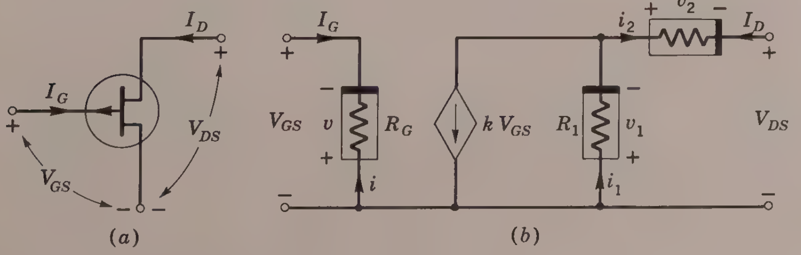


Fig. 11-23. A *p*-channel FET model A.

Exercise 3: Repeat Exercise 1 for the *n-p-n* transistor type 2NXX15.

Exercise 4: Repeat Exercise 1 for the *n-p-n* transistor type 2NXX17A.

Exercise 5: Repeat Exercise 1 for the *p-n-p* transistor type 2NXX16.

Exercise 6: Repeat Exercise 1 for the *p-n-p* transistor type 2NXX18A.

Exercise 7: (*a*) Find all pertinent parameters associated with the *n*-channel FET model A for the FET type 2NXX19. (*b*) Sketch the characteristic curves predicted by this model and compare with the actual curves.

Exercise 8: Repeat Exercise 7 for the FET type 2NXX20.

Exercise 9: Explain why the models for the transistors in the common-base configuration are different from those in the common-emitter configuration; i.e., they cannot be obtained by simply interchanging the base and the emitter terminals. How can we reconcile this with our earlier observation that the characteristics corresponding to different reference terminals are equivalent, and can be derived from one another?

11-7 SUMMARY

Classification of models A model can be either a mathematical model or a network model. Mathematical models are useful for numerical analysis, while network models are useful for qualitative and quantitative circuit analysis. For convenience, network models usually contain piecewise-linear resistors and linear controlled sources. They are ideally suited for the iterative piecewise-linear method to be presented in the next chapter.

Types of network models A network model may be either a global model or a local model. A global model must approximate the characteristic curves of the device over the entire v-i plane. A local model need approximate the curves over only some prescribed region of the v-i plane.

Basic philosophy of network model making A model must always be a compromise between simplicity and reality. Some networks may require a global model; others may require only a local model. If the region of operation is known, as from previous experience, then a local model should be chosen. This is because the local model (*extracted* from a global model) is always simpler. We can afford, therefore, to increase the accuracy of local models by adding "frosting elements" without increasing their complexity to unmanageable proportions.

Modeling controlled resistors A large class of controlled resistors can be modeled by a network containing two-terminal resistors and controlled sources. When the curves are uniformly spaced and parallel with respect to some axis, the model consists of a nonlinear resistor, a linear controlled source, and possibly a rotator. When the spacing is not uniform, a nonlinear controlled source may be needed.

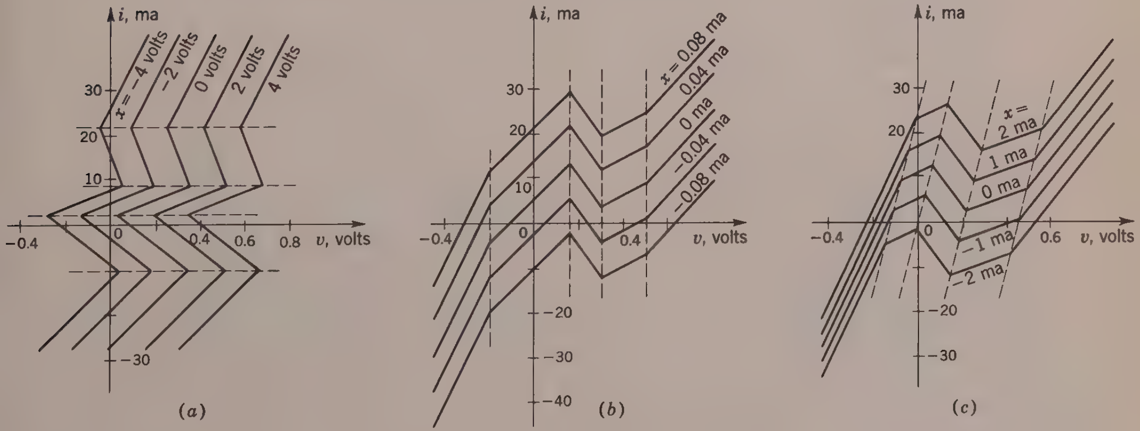
Modeling three-terminal resistors Any three-terminal resistor can be modeled by an interconnection of two controlled resistors. In many cases, a simpler realization can be obtained by choosing two basic models:

1. The T-network model
2. The π-network model

PROBLEMS

11-1 Synthesize a global network model to represent each set of controlled-resistor characteristic curves shown in Fig. P-11-1*a* to *c*.

Fig. P11-1.



(*a*) (*b*) (*c*)

11-2 Synthesize a global network model to represent the three-terminal resistor characteristic curves shown in Fig. P11-2.

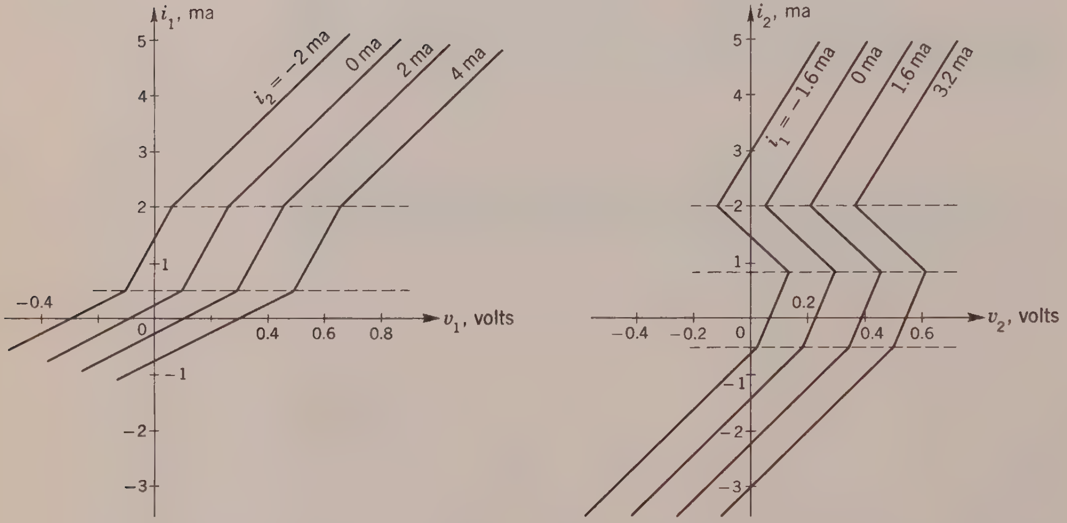
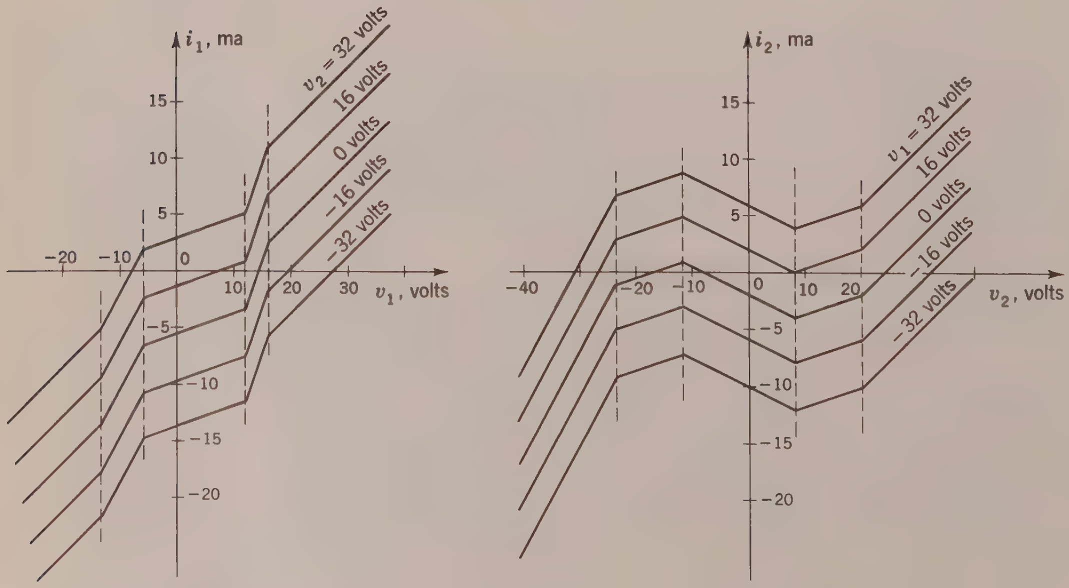


Fig. P11-2.

11-3 Synthesize a global network model to represent the three-terminal resistor characteristic curves shown in Fig. P11-3.

Fig. P11-3.



11-4 Synthesize a global network model to represent the three-terminal resistor characteristic curves shown in Fig. P11-4.

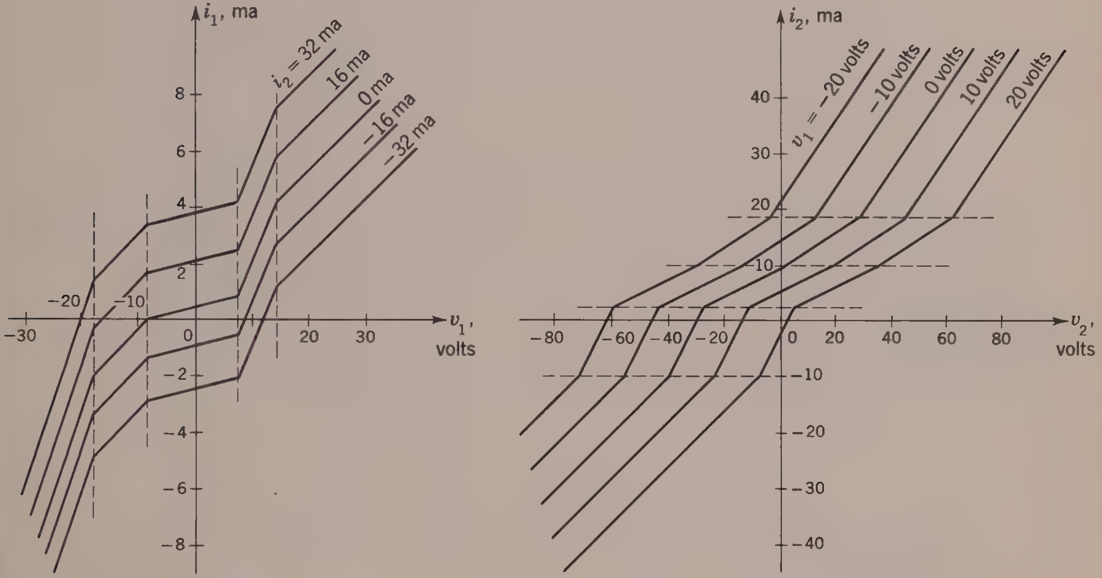
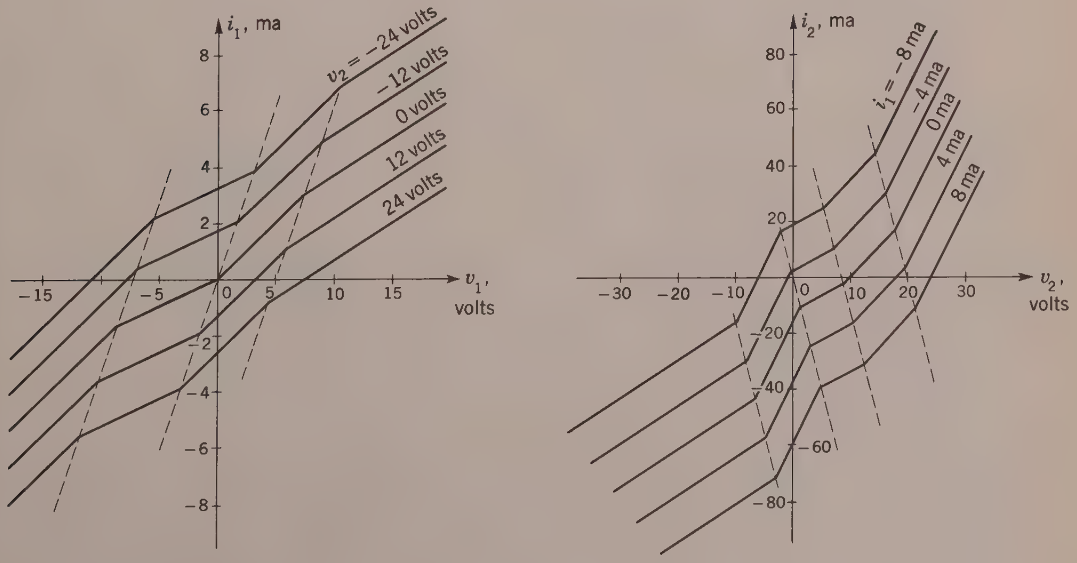


Fig. P11-4.

11-5 Synthesize a global network model to represent the three-terminal resistor characteristic curves shown in Fig. P11-5.

Fig. P11-5.



11-6 Synthesize a global network model to represent the three-terminal resistor characteristic curves shown in Fig. P11-6.

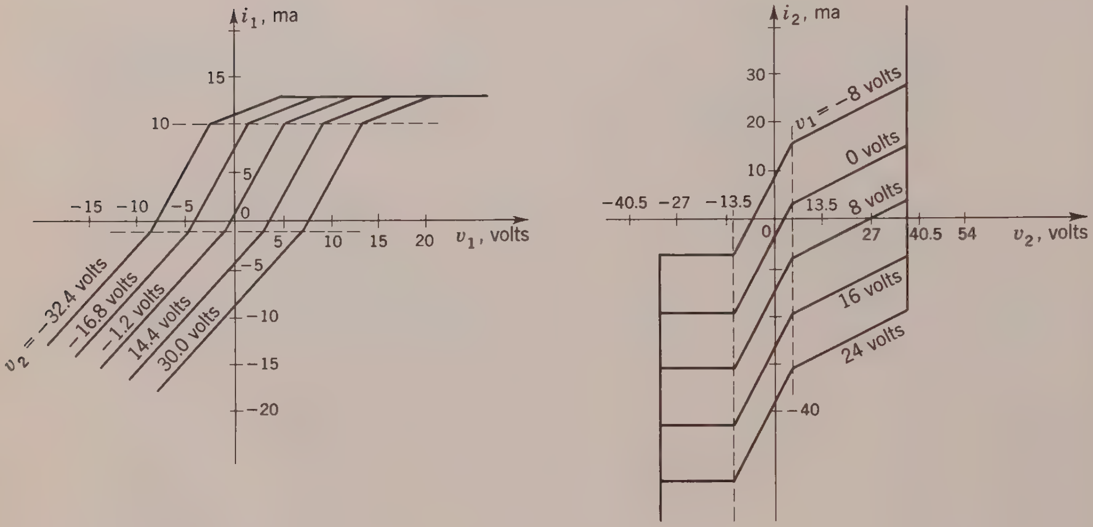
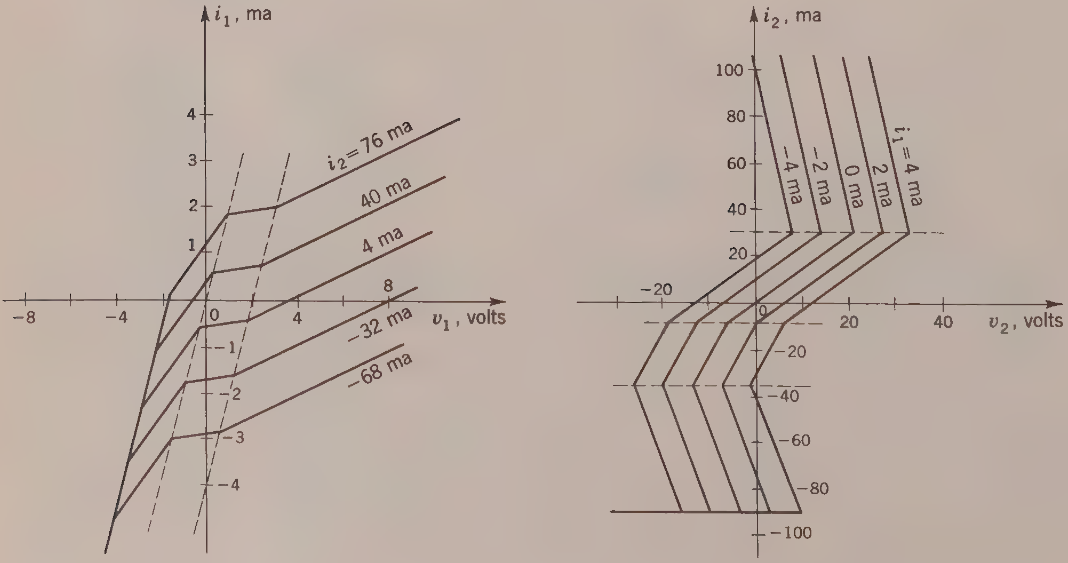


Fig. P11-6.

11-7 Synthesize a global network model to represent the three-terminal resistor characteristic curves shown in Fig. P11-7.

Fig. P11-7.



11-8 Synthesize a global network model for the vacuum tetrode type 2NXX23.

11-9 Synthesize a global network model for the unijunction transistor type 2NXX21.

11-10 Synthesize a global network model for the SCR type 2NXX14.

11-11 Synthesize a global network model for the NTC thermistor type 1NXX12.

11-12 Synthesize a global network model for the PTC thermistor type 1NXX13.

11-13 Modify the global network model obtained in Prob. 11-3 so that the resulting characteristic curves are constrained to the first and the third quadrants only.

11-14 Modify the global network model obtained in Prob. 11-4 so that the resulting characteristic curves are constrained to the first and the third quadrants only.

12 ITERATIVE PIECEWISE-LINEAR ANALYSIS AND SYNTHESIS OF RESISTIVE NONLINEAR NETWORKS

12-1 BASIC PHILOSOPHY OF THE ITERATIVE PIECEWISE-LINEAR METHOD

The methods of analysis and synthesis presented in the preceding chapters are graphical in nature. These methods are simple and are applicable to many practical nonlinear resistive network problems. Unfortunately, these graphical methods are not general enough to handle the more complicated non-series-parallel networks containing one or more nonlinear three-terminal resistors. Neither are they applicable to nonlinear networks containing controlled sources. Because of these limitations, it is necessary that more general methods of analysis be found to supplement the graphical methods. So far, the most general method of analysis is the *iterative piecewise-linear method.*[1] This method is applicable to any nonlinear network in the *canonic form.* A network is said to be in canonic form if it satisfies the following two assumptions:

1. All nonlinear elements in the network are two-terminal resistors.
2. The v-i curve of each nonlinear resistor in the network is represented by piecewise-linear segments.

These assumptions are not as restrictive as they seem to indicate. Indeed, if a network is not already in the canonic form, we can always transform it into the canonic form. This can be done by the following steps:

Step 1 Apply the principles of Chap. 11 to substitute each three-terminal resistor or controlled resistor in the network by a network model containing only nonlinear two-terminal resistors, controlled sources, independent source, and other linear elements, such as linear resistors, rotators, ideal transformers, NICs, and gyrators.

---

[1] A more rigorous exposition of this method is given in L. O. Chua, Analysis and Synthesis of Multivalued Memoryless Nonlinear Networks, *IEEE Trans. Circuit Theory,* vol. CT-14, no. 2, pp. 192–209, June, 1967.

Step 2 Represent the v-i curve of each nonlinear resistor by piecewise-linear segments. Therefore, without any loss of generality we shall, for convenience, assume that all networks in this chapter are already in the canonic form.

The basic philosophy of the iterative piecewise-linear method consists of treating a nonlinear network as if it were a number of distinct linear networks having identical network topology. Each linear network can then be analyzed independently by well-known linear techniques. This is possible because by assumption the v-i curves of all nonlinear resistors are represented by piecewise-linear segments. If we know that corresponding to a given input voltage or current, the network is operating at a certain segment of each nonlinear resistor in the network, then we can extend the corresponding segment of each resistor beyond its breakpoint, thus forming a new linear network, without affecting the solution of the original network. We shall now render this concept more precise.

Let N be a dc nonlinear resistive network in the canonic form. Therefore N may contain any number of linear elements (positive or negative resistances, rotators, ideal transformers, NICs, gyrators, etc.), dc independent sources, controlled sources, and nonlinear two-terminal resistors characterized by arbitrary piecewise-linear v-i curves (may be multivalued or even disconnected). Let these nonlinear resistors be labeled consecutively from R_1 to R_m, where m is the total number of nonlinear resistors in the network.

Corresponding to each nonlinear resistor R_j, let s_j be the number of piecewise-linear segments of the associated v-i curve. Let ${}_kg_j = 1/{}_kr_j$ represent the slope (conductance) of segment k of R_j.† Let segment k of R_j or its extension intersect the v_j and i_j axes at $v_j = {}_kE_j$ and $i_j = {}_kI_j$, respectively. Let segment k of R_j be defined for all ${}_kE_j^- < v_j < {}_kE_j^+$ and for all ${}_kI_j^- < i_j < {}_kI_j^+$. Figure 12-1$a$ shows a typical segment k of R_j, illustrating the various symbols defined above. Observe that each segment k is completely specified by:

1. Slope of segment $k = {}_kg_j = 1/{}_kr_j$.
2. Voltage intercept of segment $k = {}_kE_j$ or current intercept of segment $k = {}_kI_j$.
3. Interval of definition for segment k, namely, ${}_kE_j^- < v_j < {}_kE_j^+$ or ${}_kI_j^- < i_j < {}_kI_j^+$. We shall denote this by ${}_kD_j(v) = ({}_kE_j^-, {}_kE_j^+)$ or

---

† Throughout this chapter, the right subscript j will always refer to the nonlinear resistor R_j and the left subscript k will refer to segment k of the v_j-i_j curve of R_j.

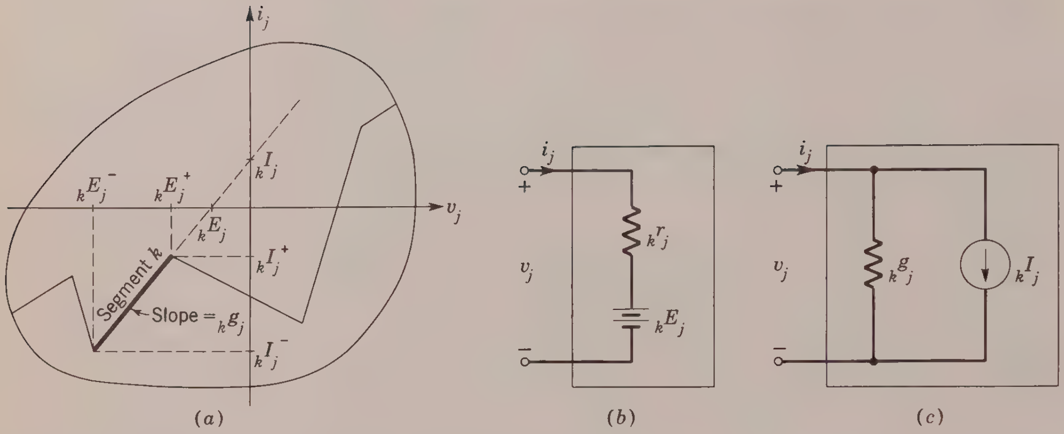


Fig. 12-1. The parameters defining a piecewise-linear segment and the associated iterative Thévenin or Norton equivalent circuits.

${}_kD_j(i) = ({}_kI_j^-, {}_kI_j^+)$. The variable v or i inside the parentheses indicates whether the interval of definition is along the voltage or the current axis.

In view of the above notations, the voltage axis v_j of R_j is subdivided into intervals, each defining a particular segment. Similarly, the current axis i_j of R_j is subdivided into the same number of intervals. Observe that if we restrict the terminal voltage v_j and terminal current i_j to lie within the prescribed intervals of definition ${}_kD_j(v)$ and ${}_kD_j(i)$ for segment k, then R_j can be replaced by an equivalent linear network as shown in Fig. 12-1*b* or *c*. By an appropriate choice of values for the parameters ${}_kr_j$, ${}_kE_j$, ${}_kg_j$, and ${}_kI_j$, the linear equivalent networks shown in Fig. 12-1*b* and *c* can represent any segment of the v_j-i_j curve. Clearly, there are as many sets of parameter values for ${}_kr_j$, ${}_kE_j$, ${}_kg_j$, and ${}_kI_j$ as there are segments. Since the equivalent networks of Fig.12-1*b* and *c* are simply the Thévenin and Norton equivalent circuit for segment j of R_j, and since there are s_j distinct sets of parameters to iterate, we shall call the equivalent networks of Fig. 12-1*b* and *c* the iterative Thévenin equivalent network and the iterative Norton equivalent network, respectively.

As an example, suppose the v_j-i_j curve of R_j is as shown in Fig. 12-2*a*. Let us tabulate the important parameters of this curve as shown in Fig. 12-2*b*. Notice that this table contains more information than is necessary to plot the v_j-i_j curve. Indeed, only columns 2, 4, and 6 or columns 3, 5, and 7 are needed to draw the v_j-i_j curve. Observe that some entries in the table may assume the value zero or infinity. Observe also that there are some entries which are undefined, namely, the intervals ${}_5D_j(v)$ and ${}_6D_j(i)$. These special cases will occur whenever the pertinent segment is either vertical or horizontal. It is, therefore, necessary to choose the iterative Thévenin equivalent circuit to completely define a vertical segment and the iterative Norton equivalent circuit to completely define a horizontal segment.

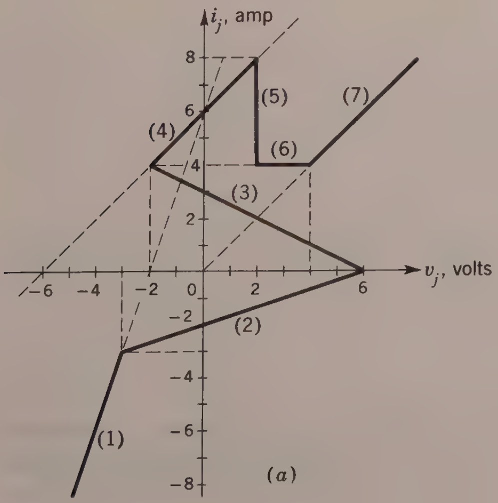
It should now be clear that if we replace each nonlinear resistor in a network N by its iterative Thévenin or Norton equivalent circuit, the result is a linear network N' having a topology identical with N. The parameters associated with the equivalent circuit of each nonlinear resistor R_j clearly depend on the particular segment under consideration. If there are m nonlinear resistors, each containing s_j segments, then there are a total of n segment combinations, where

$$n = s_1 s_2 \cdots s_j \cdots s_m \tag{12-1}$$

Since N' may assume n different sets of parameter values, we shall call it the iterative linear network of N. The iterative piecewise-linear method to be presented in the following sections is nothing more than a systematic analysis of the iterative linear network associated with a given nonlinear network.

Exercise 1: Show that if the v_j-i_j curve of a resistor R_j contains one or more horizontal (vertical) segments, it can always be converted into an equivalent v_j'-i_j' curve without horizontal (vertical) segments in parallel (series) with a linear resistor.

Fig. 12-2. An example illustrating the parameters associated with a typical v_j-i_j curve.



(a)

k	${}_kr_j$	${}_kg_j$	${}_kE_j$	${}_kI_j$	${}_kD_j(v)$	${}_kD_j(i)$
1	1/3	3	−2	6	(−∞, −3)	(−∞, −3)
2	3	1/3	6	−2	(−3, 6)	(−3, 0)
3	−2	−1/2	6	3	(−2, 6)	(0, 4)
4	1	1	−6	6	(−2, 2)	(4, 8)
5	0	∞	2	∞	—	(4, 8)
6	∞	0	∞	4	(2, 4)	—
7	1	1	0	0	(4, ∞)	(4, ∞)

(b)

Exercise 2: Show that if the v_j-i_j curve of a resistor R_j contains both horizontal and vertical segments, it can always be converted into an equivalent v_j'-i_j' curve having neither horizontal nor vertical segments by using two linear resistors. HINT: See Exercise 1.

Exercise 3: Show that if a network contains n nonlinear resistors each characterized by a two-segment v_j-i_j curve, there are a total of 2^n distinct segment combinations.

12-2 DETERMINATION OF OPERATING POINTS

Before we present the iterative piecewise-linear method in its complete generality, let us consider the simple network shown in Fig. 12-3*a*. For convenience, assume that the nonlinear resistor R_1 is characterized by the multivalued v-i curve shown in Fig. 12-2*a*. We could have immediately obtained the operating points of this network graphically as shown in Fig. 12-3*b*. Instead of doing this, however, we shall determine the operating points by the iterative piecewise-linear method and check the answers with those obtained in Fig. 12-3*b*.

Consider first an arbitrary operating point, such as Q_1 of Fig. 12-3*b*. Observe that only the segment containing Q_1 is pertinent in the determination of this point. This means that we can substitute the v_1-i_1 curve by a straight line formed by segment 2 and its extensions and still obtain the same solution.[1] Hence, we can replace the resistor R_1 either by its iterative Thévenin equivalent circuit as shown in Fig. 12-3*c* or by its iterative Norton equivalent circuit as shown in Fig. 12-3*d*. We can next solve for i_1 and v_1 from Fig. 12-3*c* to obtain

$$i_1 = \frac{4 - {}_kE_1}{1 + {}_kr_1} \tag{12-2}$$

$$v_1 = ({}_kr_1)i_1 + {}_kE_1 \tag{12-3}$$

Similarly, we can solve for i_1 and v_1 from Fig. 12-3*d* to obtain

$$i_1 = \frac{4}{1 + (1/{}_kg_1)} + \frac{{}_kI_1}{1 + {}_kg_1} \tag{12-4}$$

$$v_1 = \frac{(i_1 - {}_kI_1)}{{}_kg_1} \tag{12-5}$$

Observe that the above expressions for i_1 and v_1 are functions of the parameters ${}_kr_1$, ${}_kE_1$, ${}_kg_1$, and ${}_kI_1$. It is necessary to substitute appropriate values for these parameters to obtain the actual values

---

[1] The reader should recognize this as a direct application of the equivalent operating-point concept in Sec. 7-4.

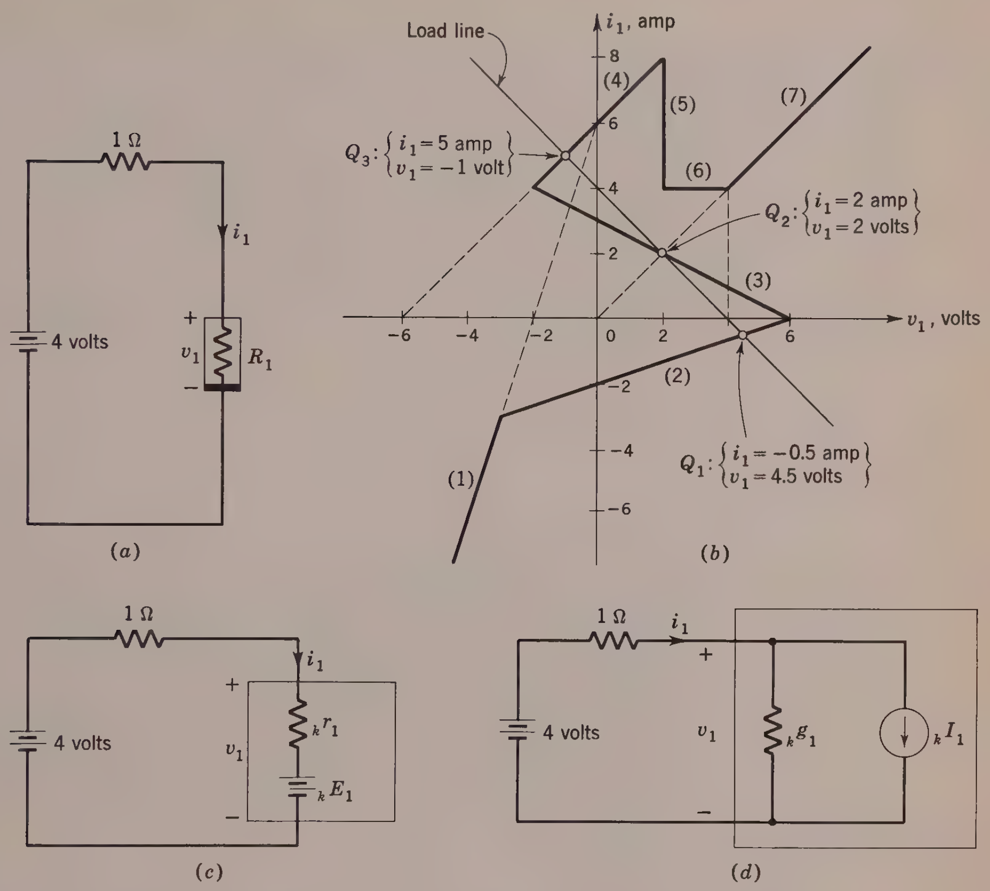


Fig. 12-3. An example of the determination of the operating points of a simple network containing a multivalued resistor.

for i_1 and v_1 at the operating point. For example, we can obtain the coordinates of Q_1 by substituting the parameters in row 2 of the table in Fig. 12-2*b* into Eqs. (12-2) and (12-3) to obtain

$$i_1 = \frac{4-6}{1+3} = -0.5 \qquad v_1 = 3(-0.5) + 6 = 4.5$$

or into Eqs. (12-4) and (12-5) to obtain

$$i_1 = \frac{4}{1+3} + \frac{-2}{1+(\frac{1}{3})} = -0.5 \qquad v_1 = \frac{-0.5-(-2)}{(\frac{1}{3})} = 4.5$$

In either case, we obtain the same answer. However, for this particular network, it is easier to use the iterative Thévenin equivalent circuit of Fig. 12-3*c*.

Observe that we have chosen the parameters pertaining to segment 2 in determining Q_1 because we happen to know from Fig. 12-3*b* that Q_1 lies on segment 2. In general, we do not know the pertinent segments associated with the operating points. However, it is not essential to know this information. We can always assume an arbitrary segment and proceed to calculate the operating point. If the result falls within the interval of definition of the assumed segment, then it is a valid operating point. For example, our assumption for segment 2 yields the point Q_1, which falls within the intervals of definition ${}_2D_1(i) = (-3,0)$ and ${}_2D_1(v) = (-3,6)$. We shall denote this fact by the standard mathematical notation

$$i_1 \in {}_2D_1(i) \tag{12-6}$$

and

$$v_1 \in {}_2D_1(v) \tag{12-7}$$

On the other hand, if the resulting operating point falls outside the interval of definition, then we conclude that our assumption is incorrect and the operating point is inconsistent. For example, had we substituted the parameters pertaining to segment 1 in Eqs. (12-2) and (12-3), we would obtain $i_1 = 4.5$ and $v_1 = -0.5$. Since the intervals of definition for segment 1 are ${}_1D_1(i) = (-\infty,-3)$ and ${}_1D_1(v) = (-\infty,-3)$, this operating point falls outside segment 1 and is, therefore, inconsistent. We shall denote the inconsistency by

$$i_1 \notin {}_1D_1(i) \tag{12-8}$$

$$v_1 \notin {}_1D_1(v) \tag{12-9}$$

If we iterate the above procedure over all possible segments of the v_1-i_1 curve and discard all operating points which are inconsistent, we would obtain the desired set of operating points. Table 12-1 shows a sample solution for the network of Fig. 12-3*a*. Observe that the values of both i_1 and v_1 are calculated and listed in this table. In practice, however, as soon as i_1 is found to be inconsistent with a certain segment, there is no need to calculate v_1, and vice versa. Instead, we can proceed immediately to the next iteration. The variables v_1 and i_1 can generally be calculated from either the iterative Thévenin or Norton equivalent circuit. Observe, however, that for segment 5 it is necessary to use the iterative Thévenin equivalent circuit, and for segment 6 it is

TABLE 12-1 *A sample solution.*

Segment	${}_kD_1(i)$	i_1 Eq. (12-2) or (12-4)	${}_kD_1(v)$	v_1 Eq. (12-3) or (12-5)	Consistency
$k = 1$	$(-\infty,-3)$	$i_1 = 4.5 \notin {}_1D_1(i)$	$(-\infty,-3)$	$v_1 = -0.5 \notin {}_1D_1(v)$	No
$k = 2$	$(-3,0)$	$i_1 = -0.5 \in {}_2D_1(i)$	$(-3,6)$	$v_1 = 4.5 \in {}_2D_1(v)$	Yes
$k = 3$	$(0,4)$	$i_1 = 2 \in {}_3D_1(i)$	$(-2,6)$	$v_1 = 2 \in {}_3D_1(v)$	Yes
$k = 4$	$(4,8)$	$i_1 = 5 \in {}_4D_1(i)$	$(-2,2)$	$v_1 = -1 \in {}_4D_1(v)$	Yes
$k = 5$	$(4,8)$	$i_1 = 2 \notin {}_5D_1(i)$	—	$v_1 = 2$	No
$k = 6$	—	$i_1 = 4$	$(2,4)$	$v_1 = 0 \notin {}_6D_1(v)$	No
$k = 7$	$(4,\infty)$	$i_1 = 2 \notin {}_7D_1(i)$	$(4,\infty)$	$v_1 = 2 \notin {}_7D_1(v)$	No

necessary to use the iterative Norton equivalent circuit. The last column of Table 12-1 indicates that of the seven iterations, only three ($k = 2,3,4$) operating points are consistent. These points, of course, coincide with those obtained in Fig. 12-3*b*.

The same philosophy can be easily extended to determine the operating points of networks containing more than one nonlinear resistor. As usual, the first step consists of substituting each of the m nonlinear resistors by its iterative Thévenin or Norton equivalent circuit. This substitution transforms the given nonlinear network N into a *linear* network N'. Therefore, we can apply any linear technique to solve for any variable of interest. The resulting solution will be a function of the iterative network parameters ${}_kr_j$, ${}_kE_j$, ${}_kg_j$, and ${}_kI_j$. However, regardless of the variables of interest, it is necessary to first solve for either the current i_j or the voltage v_j of each nonlinear resistor R_j. This is essential because these variables must eventually be checked to see whether their values are consistent with the assumptions imposed in the next step.

The next step consists of assuming a certain combination of segment numbers of the m v_j-i_j curves and substituting the parameter values associated with the respective segments into the expressions for v_j and i_j, $j = 1, 2, \ldots, m$.

The final step consists of checking the solutions corresponding to a particular segment combination and determining whether the values of v_j and i_j *simultaneously* fall within the intervals of definition of the respective segments. If this is not so, the operating point is inconsistent and must be discarded. We then proceed to the next iteration with a new segment combination. Since the network may contain more than one operating point, it is generally necessary to iterate over all possible segment combinations, that is, a total of n times, where n is as defined in Eq. (12-1).[1]

---

[1] For a more efficient method which does not require iterating over all segment combinations, see L. O. Chua, An Efficient Algorithm for Implementing the Iterative Piecewise-Linear Method, *Tenth Semi-Annual Research Summary,* School of Electrical Engineering, Purdue University, Lafayette, Ind., January–June, 1969.

EXAMPLE

1. Consider the network shown in Fig. 12-4*a*. The v_j-i_j curves of R_1 and R_2 are shown in Fig. 12-4*b* and *c*. Our first step is to replace R_1 and R_2 by their iterative Thévenin equivalent circuits as shown in Fig. 12-4*d*.[1] For ease of reference we have tabulated the pertinent parameters associated with each segment of the v_j-i_j curve as shown in Fig. 12-4*e*.

   The next step is to solve for either v_j or i_j associated with R_1 and R_2. For the circuit of Fig. 12-4*d*, it is easier to solve for i_1 and i_2. Since this network is linear, any linear technique is acceptable. The most general method, although not necessarily the simplest computationally, is to write the equations of motion of the network and solve for i_1 and i_2. The equations of motion for this network are

$$(6 + {}_kr_1)i_1 - 6i_2 = 2.5 - {}_kE_1 \tag{12-10}$$

$$-6i_1 + (7 + {}_kr_2)i_2 = -{}_kE_2 \tag{12-11}$$

   Solving for i_1 and i_2, we obtain

$$i_1 = \frac{1}{\Delta}\{(7 + {}_kr_2)2.5 - [(7 + {}_kr_2){}_kE_1 + 6({}_kE_2)]\} \tag{12-12}$$

$$i_2 = \frac{1}{\Delta}\{15 - [6({}_kE_1) + (6 + {}_kr_1){}_kE_2]\} \tag{12-13}$$

   where

$$\Delta \equiv (6 + {}_kr_1)(7 + {}_kr_2) - 36 \tag{12-14}$$

   As expected, i_1 and i_2 are functions of the parameters ${}_kr_j$ and ${}_kE_j$. Our next step is to substitute values for these parameters corresponding to each segment combination. Since $m = 2$, $s_1 = 2$, and $s_2 = 2$, we have to iterate $n = 2(2) = 4$ times. For simplicity let us denote the combination "segment k_1 of R_1 and segment k_2 of R_2" by the notation "segment (k_1,k_2)." For example, segment (1,1) denotes segment 1 of R_1 and segment 1 of R_2. We are now ready to perform the following iteration:

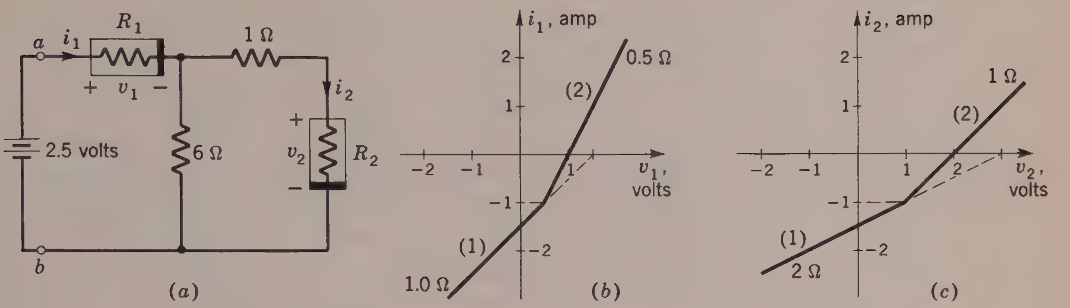
   Segment (1,1): $\Delta = (6 + 1)(7 + 2) - 36 = 27$

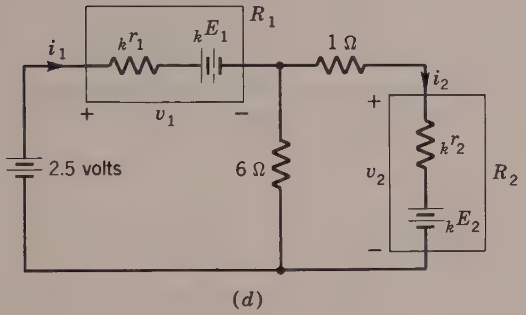
$$i_1 = \tfrac{1}{27}\{(7 + 2)(2.5) - [(7 + 2)(1.5) + 6(3)]\} = -\tfrac{1}{3} \text{ amp}$$

   Since ${}_1D_1(i) = (-\infty, -1)$, $i_1 \notin {}_1D_1(i)$, and hence this solution is inconsistent. There is no need to proceed and calculate i_2 since even if i_2 is consistent, the resulting operating point must still be discarded. Let us, therefore, perform our next iteration.

---

[1] Alternately, we can replace R_1 and R_2 by their iterative Norton equivalent circuit. We can also replace R_1 by its iterative Thévenin equivalent circuit and R_2 by its iterative Norton equivalent circuit, or vice versa. A careful choice of the equivalent circuit will sometimes simplify the linear analysis part of the procedure.





R_j	k	$_kr_j$	$_kE_j$	$_kD_j(v)$	$_kD_j(i)$
R_1	1	1.0	1.5	$(-\infty, 0.5)$	$(-\infty, -1)$
	2	0.5	1.0	$(0.5, \infty)$	$(-1, \infty)$
R_2	1	2.0	3.0	$(-\infty, 1)$	$(-\infty, -1)$
	2	1.0	2.0	$(1, \infty)$	$(-1, \infty)$

(e)

Fig. 12-4. The nonlinear network associated with Example 1 and its iterative piecewise-linear equivalent circuit.

We have

Segment (1,2): $\Delta = (6 + 1)(7 + 1) - 36 = 20$

$$i_1 = \tfrac{1}{20}\{(7 + 1)(2.5) - [(7 + 1)(1.5) + 6(2)]\} = -\tfrac{1}{5} \text{ amp}$$

Again, $i_1 \notin {}_1D_1(i)$, so we proceed to the next iteration, obtaining

Segment (2,1): $\Delta = (6 + 0.5)(7 + 2) - 36 = 22.5$

$$i_1 = \tfrac{1}{22.5}\{(7 + 2)(2.5) - [(7 + 2)(1) + 6(3)]\} = -\tfrac{1}{5} \text{ amp}$$

This time we find $i_1 \in {}_2D_1(i)$, and the solution for i_1 is consistent. Observe, however, that we cannot conclude that the operating point corresponding to this segment is consistent until we have checked i_2. In fact, we found

$$i_2 = \tfrac{1}{22.5}\{15 - [6(1) + (6 + 0.5)(3)]\} = -\tfrac{7}{15} \text{ amp}$$

and $i_2 \notin {}_1D_2(i)$. Therefore, this operating point is also inconsistent. Let us proceed to make our last iteration and obtain

Segment (2,2): $\Delta = (6 + 0.5)(7 + 1) - 36 = 16$

$$i_1 = \tfrac{1}{16}\{(7 + 1)(2.5) - [(7 + 1)(1.0) + 6(2)]\} = 0 \text{ amp}$$

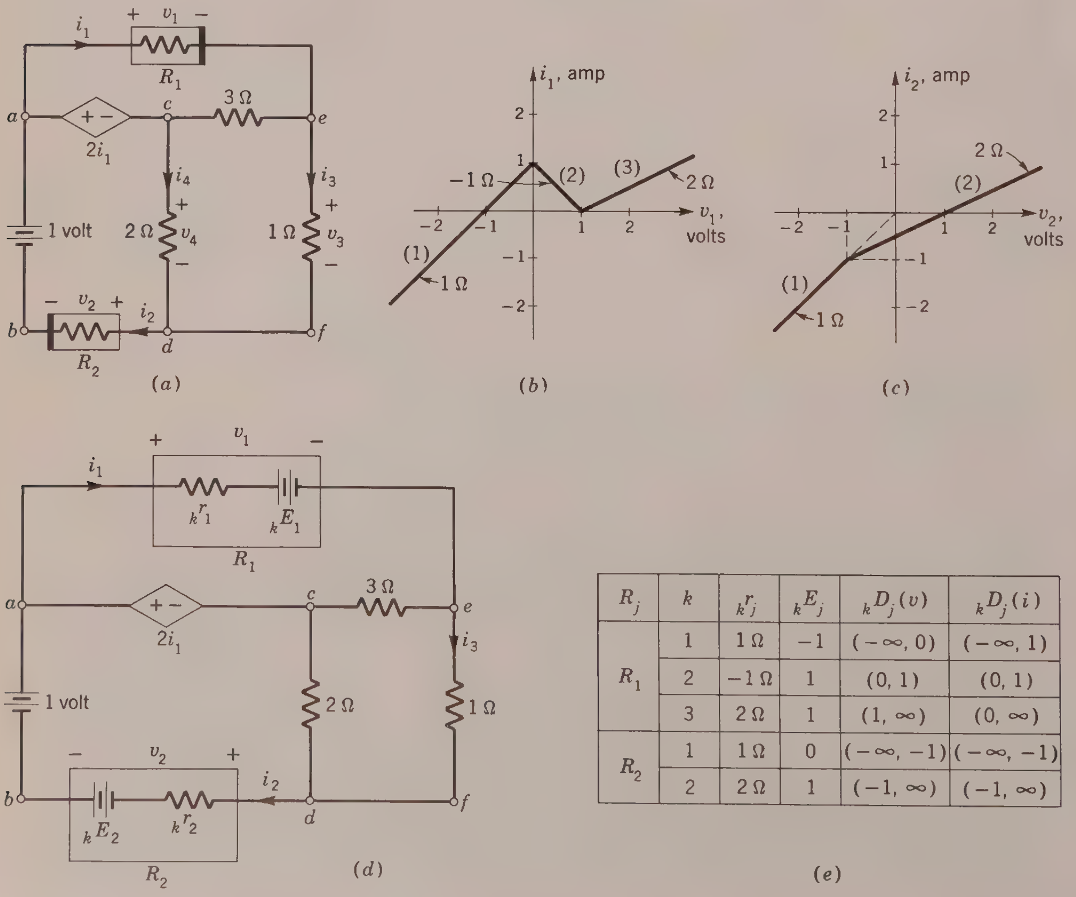
Since $i_1 \in {}_2D_1(i)$, it is consistent, and we proceed to calculate i_2 as follows:

$$i_2 = \tfrac{1}{16}\{15 - [6(1) + (6 + 0.5)(2)]\} = -\tfrac{1}{4} \text{ amp}$$

Since $i_2 \in {}_2D_2(i)$, it is also consistent. Out of four possibilities, we found only one consistent solution. Therefore, this network has only one operating point, that is, $i_1 = 0$ amp and $i_2 = -\tfrac{1}{4}$ amp. From these we can easily determine the values of the remaining variables from Fig. 12-4*b* and *c*; namely, $v_1 = 1$ volt and $v_2 = 1.75$ volts.

Observe that the order of the iteration is unimportant. For example, we could just as well iterate in the following order: segment (2,2), segment (1,1), segment (2,1), and segment (1,2). In this case we would obtain a consistent operating point in our first iteration. If the network is known to have only one

Fig. 12-5. The nonlinear network associated with Example 2 and its iterative piecewise-linear equivalent circuit.



R_j	k	$_kr_j$	$_kE_j$	$_kD_j(v)$	$_kD_j(i)$
R_1	1	$1\,\Omega$	-1	$(-\infty, 0)$	$(-\infty, 1)$
	2	$-1\,\Omega$	1	$(0, 1)$	$(0, 1)$
	3	$2\,\Omega$	1	$(1, \infty)$	$(0, \infty)$
R_2	1	$1\,\Omega$	0	$(-\infty, -1)$	$(-\infty, -1)$
	2	$2\,\Omega$	1	$(-1, \infty)$	$(-1, \infty)$

solution, which is the case here, in view of the uniqueness theorem in Chap. 5, we need not iterate any further. On the other hand, if we do not know beforehand the number of possible operating points, such as the network in the following examples,[1] then it will be necessary to iterate over all segment combinations.

2. The network shown in Fig. 12-5*a* contains a controlled source and two nonlinear resistors whose v_j-i_j curves are as shown in Fig. 12-5*b* and *c*. Replacing these resistors by their iterative Thévenin equivalent circuits, we obtain the iterative linear network shown in Fig. 12-5*d*. The pertinent parameters are listed in Fig. 12-5*e*.

   As usual, our first objective is to solve for i_1 and i_2 by any linear technique. The results of the linear analysis are

$$i_1 = \frac{1 - (\frac{4}{3} + {}_kr_2)\,{}_kE_1 - {}_kE_2}{\frac{1}{3} + \frac{4}{3}({}_kr_1) - \frac{1}{2}({}_kr_2) + ({}_kr_1)({}_kr_2)} \tag{12-15}$$

$$i_2 = \frac{-\frac{1}{2} + {}_kr_1 + \frac{1}{2}({}_kE_2) - ({}_kr_1)({}_kE_2) + {}_kE_1}{\frac{1}{3} + \frac{4}{3}({}_kr_1) - \frac{1}{2}({}_kr_2) + ({}_kr_1)({}_kr_2)} \tag{12-16}$$

Since $m = 2$, $s_1 = 3$, and $s_2 = 2$, we need to iterate six times as follows:

Segment (1,1) $i_1 = 1.53$ $\notin {}_1D_1(i)$	Segment (2,1) $i_1 = \frac{8}{15}$ $\in {}_2D_1(i)$ $i_2 = 0.2$ $\notin {}_1D_1(i)$	Segment (3,1) $i_1 = -\frac{8}{27}$ $\notin {}_3D_1(i)$
Segment (1,2) $i_1 = 1.25$ $\notin {}_1D_1(i)$	Segment (2,2) $i_1 = \frac{5}{6}$ $\in {}_2D_1(i)$ $i_2 = -0.25$ $\in {}_2D_2(i)$	Segment (3,2) $i_1 = -\frac{5}{9}$ $\notin {}_3D_1(i)$

Out of six possible combinations, only the operating point corresponding to segment (2,2) is consistent; that is, $i_1 = \frac{5}{6}$ amp and $i_2 = -0.25$ amp. From these we immediately obtain $v_1 = 0.167$ volt, $v_2 = 0.5$ volt, $i_3 = \frac{1}{3}$ amp, $v_3 = \frac{1}{3}$ volt, $i_4 = -\frac{7}{12}$ amp, and $v_4 = -\frac{7}{6}$ volt.

This example shows that the iterative piecewise-linear method is applicable to networks containing controlled sources. To show that this method is also applicable to networks containing linear *n*-port elements such as ideal transformers, gyrators, and rotators, consider the following example.

---

[1] The following two examples, as well as Examples 2 and 3 in Secs. 12-3 and 12-4, involve a certain amount of tedious numerical calculations and may be omitted on first reading without impairing the understanding of the method. These examples are presented to bring out two aspects of the method: (1) Unlike other methods, the iterative piecewise-linear approach is completely general. In particular, it is applicable to networks containing controlled sources, ideal transformers, NICs, gyrators, etc. Moreover, it can handle networks having *multivalued* operating points, DP plots, and TC plots. (2) Except for the simplest nonlinear networks, the method should obviously be carried out on a computer in order to be practical.

3. The network shown in Fig. 12-6*a* contains an ideal transformer, a controlled source, and two nonlinear resistors whose v_j-i_j curves are as shown in Fig. 12-6*b* and *c*. The associated iterative linear network is shown in Fig. 12-6*d*, and the pertinent parameters are listed in Fig. 12-6*e*.

   The first step is to analyze the iterative linear network shown in Fig. 12-6*d*. The results are as follows:

$$i_1 = \frac{5 - 4({}_kE_1) - 2({}_kE_2)}{4({}_kr_1) - 2({}_kr_2) - 3} \qquad (12\text{-}17)$$

$$i_2 = -i_1 \qquad (12\text{-}18)$$

   Since $m = 2$, $s_1 = 2$, $s_2 = 2$, $n = 4$. Therefore, we must iterate four times.

Segment (1,1)	Segment (1,2)	Segment (2,1)	Segment (2,2)
$i_1 = -5/3$ $\in {}_1D_1(i)$ $i_2 = 5/3$ $\notin {}_1D_2(i)$	$i_1 = -1$ $\in {}_1D_1(i)$ $i_2 = 1$ $\in {}_2D_2(i)$	$i_1 = -15$ $\notin {}_2D_1(i)$	$i_1 = -3$ $\notin {}_2D_1(i)$

   The network has, therefore, only one operating point: $i_1 = -1$ amp and $i_2 = 1$ amp. From these, we obtain $v_1 = -1$ volt, $v_2 = 3$ volts, $v_3 = 2.25$ volts, $v_4 = 4.5$ volts, and $i_4 = 0.5$ amp.

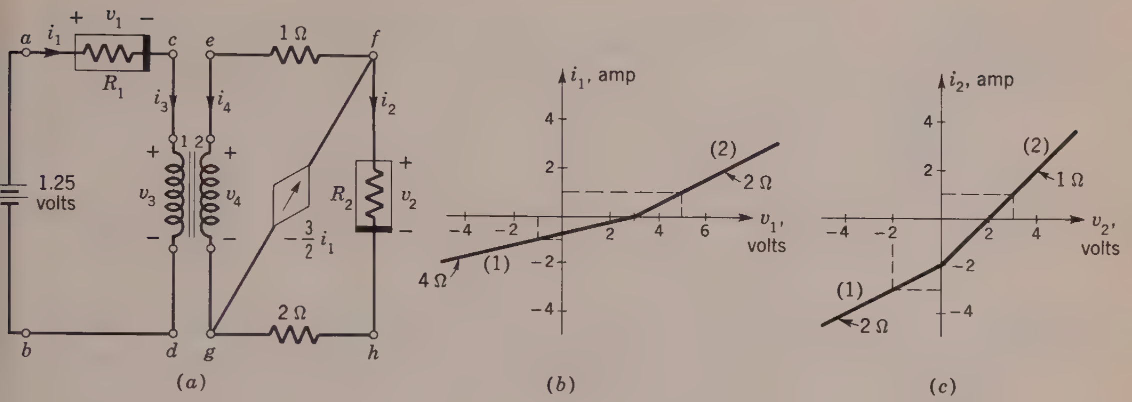
From the above examples we can see that the iterative piecewise-linear method is restricted neither by network topology nor by the complexity of the v_j-i_j curves. Indeed, it is applicable to any nonlinear network in the canonic form. The procedure can be divided into two parts: the first part consists of analyzing a linear network, and the second part consists of an iterative procedure. For networks containing more than two nonlinear resistors, the iterative part can be very tedious and time-consuming. However, since nothing more than simple substitution and comparison is involved in the iterative process, a digital computer is ideally suited for performing the iteration.

Exercise 1: Verify the operating points obtained in Example 1 by graphical methods.

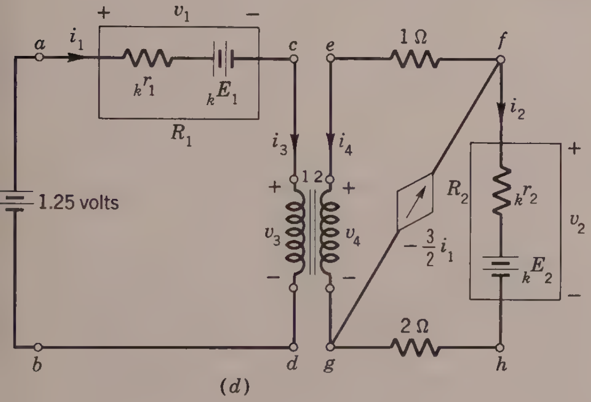
Exercise 2: Derive Eqs. (12-15) and (12-16) of Example 2.

Exercise 3: Derive Eqs. (12-17) and (12-18) of Example 3.

Exercise 4: Give a flow chart and algorithm for implementing the iteration part of the method by a computer.



(a) (b) (c)



(d)

R_j	k	${}_kr_j$	${}_kE_j$	${}_kD_j(v)$	${}_kD_j(i)$
R_1	1	4 Ω	3	$(-\infty, 3)$	$(-\infty, 0)$
	2	2 Ω	3	$(3, \infty)$	$(0, \infty)$
R_2	1	2 Ω	4	$(-\infty, 0)$	$(-\infty, -2)$
	2	1 Ω	2	$(0, \infty)$	$(-2, \infty)$

(e)

Fig. 12-6. The nonlinear network associated with Example 3 and its iterative piecewise-linear equivalent circuit.

12-3 DETERMINATION OF DP PLOTS

Before we show how the iterative piecewise-linear method can be extended to determine the DP plot of a nonlinear network, let us consider again the simple network of Fig. 12-3*a* and redraw it as shown in Fig. 12-7*a* by replacing the battery with a pair of driving-point terminals. For this simple network the DP plot can be determined graphically, and the result is given by the solid piecewise-linear curve shown in Fig. 12-7*b*.

The first step in the iterative piecewise-linear method is to draw the iterative-linear network shown in Fig. 12-7*c*. Since this network is linear, its DP plot is a straight line and can be easily obtained by deriving the relationship between v and i.

$$v = (1 + {}_kr_1)i + {}_kE_1 \qquad (12\text{-}19)$$

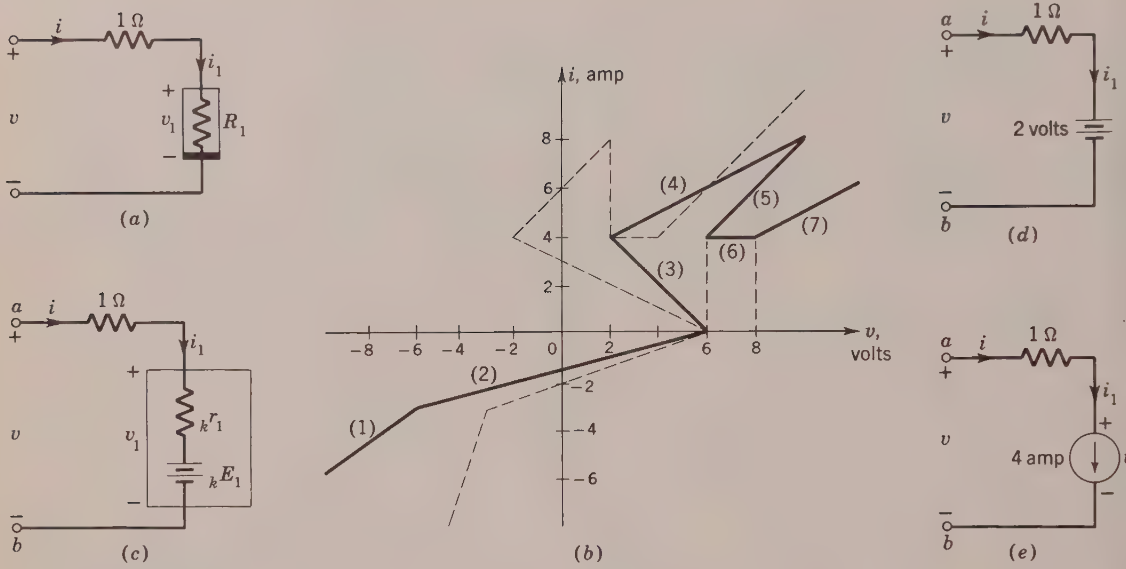


Fig. 12-7. An example of the determination of the DP plot of a simple network containing a multivalued resistor.

Corresponding to each set of parameter values for $_kr_1$ and $_kE_1$, Eq. (12-19) represents a straight line in the v-i plane. However, only a portion of this straight line corresponding to a certain range of the driving-point voltage v is valid. This is because beyond this range of v, the operating point moves to another segment of R_1, and a new set of parameter values must be substituted for $_kr_1$ and $_kE_1$ in Eq. (12-19). This will lead to another straight line in the v-i plane with a new interval of definition.

If we substitute the parameter values corresponding to all seven segments of R_1 into Eq. (12-19), we obtain seven straight lines, each containing a segment of the DP plot obtained earlier in Fig. 12-7b by graphical methods. This means that if we can find the range of values (a_k,b_k) of the driving-point voltage v for which the operating points will remain on segment k of R_1, then the DP plot of the network corresponding to segment k is determined by that portion of the corresponding straight line which lies within (a_k,b_k). The determination of the pertinent portion of each straight line defining the DP plot of the iterative linear network constitutes the crucial step in the procedure.

Since the interval of definition of each segment of the DP plot depends on the interval of definition of the corresponding segment of the v_1-i_1 curve of R_1, it is necessary to derive the relationship between either i_1 and v or between v_1 and v from the iterative linear network. By straightforward linear techniques we obtain the following two linear equations:

$$i_1 = \frac{1}{1 + {}_kr_1} v - \frac{{}_kE_1}{1 + {}_kr_1} \tag{12-20}$$

$$v_1 = \frac{{}_kr_1}{1 + {}_kr_1} v + \frac{{}_kE_1}{1 + {}_kr_1} \tag{12-21}$$

Now suppose we want to find the range of v for which the operating point will remain on segment 2 of R_1. From the v_1-i_1 curve of R_1 we obtain ${}_2r_1 = 3$, ${}_2E_1 = 6$, ${}_2D_1(i) = (-3,0)$, and ${}_2D_1(v) = (-3,6)$. Observe that segment 2 is defined in the interval

$$-3 < i_1 < 0 \tag{12-22}$$

$$-3 < v_1 < 6 \tag{12-23}$$

But i_1 and v_1 are related to v by Eqs. (12-20) and (12-21), respectively. Hence, substituting Eqs. (12-20) and (12-21) for i_1 and v_1 in Eqs. (12-22) and (12-23), we obtain

$$-3 < \frac{1}{1 + {}_kr_1} v - \frac{{}_kE_1}{1 + {}_kr_1} < 0 \tag{12-24}$$

$$-3 < \frac{{}_kr_1}{1 + {}_kr_1} v + \frac{{}_kE_1}{1 + {}_kr_1} < 6 \tag{12-25}$$

Since Eqs. (12-24) and (12-25) pertain only to segment 2, we can substitute ${}_kr_1 = {}_2r_1 = 3$ and ${}_kE_1 = {}_2E_1 = 6$ and obtain

$$-3 < \tfrac{1}{4}v - \tfrac{3}{2} < 0 \tag{12-26}$$

$$-3 < \tfrac{3}{4}v + \tfrac{3}{2} < 6 \tag{12-27}$$

Solving for v, we obtain in both cases

$$-6 < v < 6 \tag{12-28}$$

Equation (12-28) gives the interval of definition of the driving-point variable v for which the network is operating on segment 2 of R_1. Observe that either Eq. (12-20) or (12-21) is sufficient to determine this information. In fact, it is not necessary to derive i_1 or v_1 in terms of v; we could just as well derive i_1 or v_1 in terms of the driving-point current i; that is,

$$i_1 = i \tag{12-29}$$

$$v_1 = ({}_kr_1)i + {}_kE_1 \tag{12-30}$$

Substituting Eqs. (12-22) and (12-23) for i_1 and v_1 in Eqs. (12-29) and (12-30), we obtain

$$-3 < i < 0 \tag{12-31}$$

$$-3 < 3i + 6 < 6 \tag{12-32}$$

Solving for i, we obtain in both cases

$$-3 < i < 0 \tag{12-33}$$

Again, either Eq. (12-29) or (12-30) is sufficient to determine the interval of definition in terms of the driving-point current i.

As a summary, we have shown that there are generally four ways to determine the interval of definition of each segment of the DP plot; that is, we can solve for i_1 or v_1 of R_1 in terms of i or v of the driving-point terminals. In all cases, the equation represents a straight line of the form

$$x_1 = ({}_kH_1)x + {}_kK_1 \tag{12-34}$$

where $x_1 = i_1$ or v_1
$x = i$ or v
${}_kH_1$ and ${}_kK_1$ = functions of parameters ${}_kr_1$ and ${}_kE_1$

If we denote the interval of definition for segment k of R_1 by

$${}_kX_1^- < x_1 < {}_kX_1^+ \tag{12-35}$$

then we can substitute Eq. (12-34) for x_1 in Eq. (12-35) and obtain for the case where ${}_kH_1 > 0$ the inequality

$$\frac{{}_kX_1^- - {}_kK_1}{{}_kH_1} < x < \frac{{}_kX_1^+ - {}_kK_1}{{}_kH_1} \tag{12-36}$$

If ${}_kH_1 < 0$, Eq. (12-36) is still valid provided the inequality signs are reversed. This equation combines in one step the various substitutions and calculations necessary to determine the interval of definition of each segment of the DP plot of Fig. 12-7*a*. Suppose we have chosen to solve for v_1 in terms of v and have obtained Eq. (12-21). Then Eq. (12-36) becomes[1]

---

[1] Recall that $({}_kE_1^-, {}_kE_1^+)$ denotes the interval of definition of segment k of R_1. See Fig. 12-1*a*.

$$\frac{{}_kE_1^- - [{}_kE_1/(1 + {}_kr_1)]}{{}_kr_1/(1 + {}_kr_1)} < v < \frac{{}_kE_1^+ - [{}_kE_1/(1 + {}_kr_1)]}{{}_kr_1/(1 + {}_kr_1)} \qquad (12\text{-}37)$$

We can now substitute the parameters associated with each segment of R_1 to find the corresponding interval of definition from Eq. (12-37). The corresponding DP plot is obtained by substituting these parameters in Eq. (12-19). The result is tabulated in Table 12-2.

Some complication arises in substituting the parameters for segments 5 and 6 into Eq. (12-36) because segment 5 is vertical and segment 6 is horizontal. It is usually easier to consider such special cases separately.

For segment 5 we go directly to the iterative linear network shown in Fig. 12-7*d*. In this case ${}_5D_1(v)$ is undefined, and hence it is necessary to solve for i_1.

$$i_1 = v - 2 \qquad (12\text{-}38)$$

Substituting Eq. (12-38) into (see Fig. 12-2*b*)

$$4 < i_1 < 8 \qquad (12\text{-}39)$$

we obtain

$$6 < v < 10 \qquad (12\text{-}40)$$

The associated DP plot can be obtained directly from Fig. 12-7*d*.

$$v = i + 2 \qquad (12\text{-}41)$$

TABLE 12-2 *A sample DP plot solution.*

Segment	Interval of definition Eq. (12-37)	Equation of DP plot Eq. (12-19)
1	$-\infty < v < -6$	$v = (4/3)i - 2$
2	$-6 < v < 6$	$v = 4i + 6$
3	$2 < v < 6$	$v = -i + 6$
4	$2 < v < 10$	$v = 2i - 6$
5	$6 < v < 10$	$v = i + 2$
6	$6 < v < 8$	$i = 4$
7	$8 < v < \infty$	$v = 2i$

For segment 6 we go directly to the iterative linear network shown in Fig. 12-7*e*. In this case ${}_6D_1(i)$ is undefined, and hence it is necessary to solve for v_1.

$$v_1 = v - 4 \tag{12-42}$$

Substituting Eq. (12-42) into (see Fig. 12-2*b*)

$$2 < v_1 < 4 \tag{12-43}$$

we obtain

$$6 < v < 8 \tag{12-44}$$

The associated DP plot can be obtained directly from Fig. 12-8*b*.

$$i = 4 \tag{12-45}$$

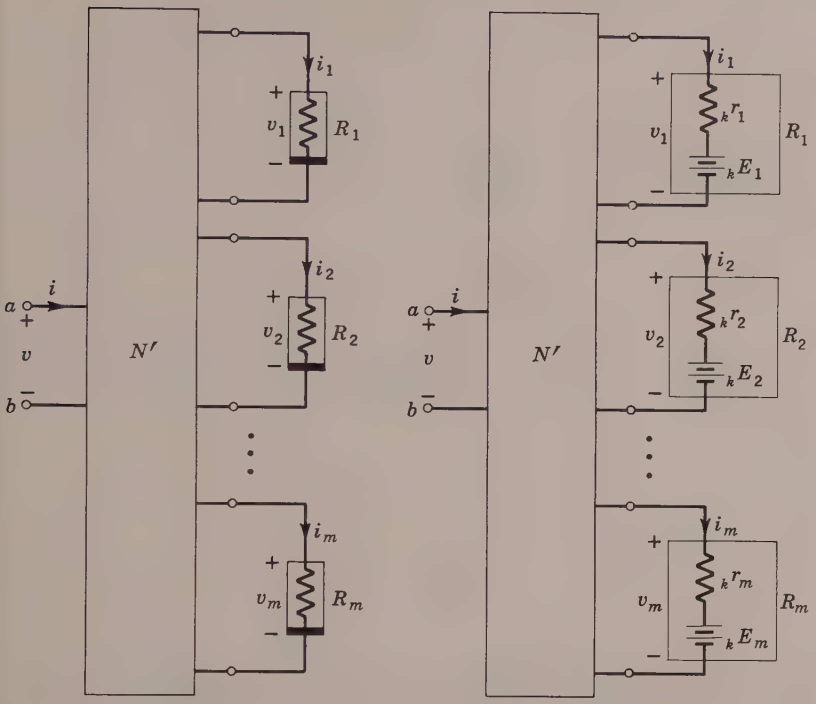
If we plot the data in Table 12-2 in the v-i plane, we would obtain the same DP plot shown in Fig. 12-7*b*.

Let us now extend the iterative piecewise-linear method to an arbitrary network N containing m nonlinear resistors, where $m > 1$. It is convenient to represent N symbolically as shown in Fig. 12-8*a*, where the m nonlinear resistors are explicitly drawn outside the remaining linear network N'. Our first step is to replace each nonlinear resistor R_j by either its iterative Thévenin or Norton equivalent circuit. To avoid unnecessary details, let us arbitrarily choose the iterative Thévenin equivalent circuit and obtain the associated iterative linear network shown in Fig. 12-8*b*. Since this network is linear, we can apply any linear technique to solve for the equation relating the driving-point voltage v and the driving-point current i. This equation can always be written in the form

$$v = Ri + E \tag{12-46}$$

where the two constants R and E are functions of the network parameter ${}_kr_j$ and ${}_kE_j$, $j = 1, 2, \ldots, m$. We have already seen an example of this in Eq. (12-19), where $m = 1$.

For each combination of parameter values, Eq. (12-46) represents a straight line in the v-i plane. Since according to Eq. (12-1) there are a total of n segment combinations, Eq. (12-46) actually represents n distinct straight lines. In order to keep track of these, it is convenient to let "segment $(k_1,k_2, \ldots ,k_m)$" denote the combi-



(a) (b)

Fig. 12-8. An arbitrary nonlinear network and its iterative piecewise-linear equivalent circuit.

nation consisting of segments k_1 of R_1, k_2 of R_2, . . . , and k_m of R_m. For example, segment (3, 5, 2, 6) denotes segments 3 of R_1, 5 of R_2, 2 of R_3, and 6 of R_4.

Since the DP plot is composed of portions of the n straight lines, the crucial step in the iterative piecewise-linear method is to determine the interval of definition of each segment combination. This can be found by the same method described earlier for the network of Fig. 12-7*a*. The only modification being that since there are now m nonlinear resistors, the procedure must be applied to each of them. This consists of deriving the relationship between each current i_j or voltage v_j in terms of the driving-point current i or the driving-point voltage v. Hence, instead of only one equation, we obtain m equations having the same form as Eq. (12-34), that is,

$$x_j = ({}_kH_j)x + {}_kK_j \qquad j = 1, 2, \ldots, m \tag{12-47}$$

where $x_j = i_j$ or v_j
$x = i$ or v
${}_kH_j$ and ${}_kK_j$ = functions of parameters ${}_kr_j$ and ${}_kE_j$

If we denote the interval of definition for segment k of R_j by

$$ {}_kX_j^- < x_j < {}_kX_j^+ \qquad j = 1, 2, \ldots, m \tag{12-48}$$

then we can substitute Eq. (12-47) for x_j in Eq. (12-48) and obtain for the case where ${}_kH_j > 0$ the inequality

$$\frac{{}_kX_j^- - {}_kK_j}{{}_kH_j} < x < \frac{{}_kX_j^+ - {}_kK_j}{{}_kH_j} \cdot \qquad j = 1, 2, \ldots, m \tag{12-49}$$

Substituting the parameter values corresponding to any segment $(k_1,k_2, \ldots ,k_m)$ into Eq. (12-49), we obtain m distinct intervals of values of the driving-point variable x; that is,

$$\begin{array}{lll} a_1 < x < b_1 & \text{or} & x \in (a_1,b_1) \\ a_2 < x < b_2 & \text{or} & x \in (a_2,b_2) \\ \cdots\cdots & & \cdots\cdots \\ a_m < x < b_m & \text{or} & x \in (a_m,b_m) \end{array} \tag{12-50}$$

These intervals must be satisfied *simultaneously* in order for the operating points to remain on segments k_1 of R_1, k_2 of $R_2, \ldots$, and k_m of R_m. This means that if we denote the interval of definition of the DP plot corresponding to segment $(k_1,k_2, \ldots ,k_m)$ by

$$D(k_1,k_2, \ldots ,k_m) \tag{12-51}$$

then $D(k_1,k_2, \ldots ,k_m)$ must be given by the largest interval common to all m intervals (a_j,b_j) of Eq. (12-50), namely,

$$D(k_1,k_2, \ldots ,k_m) = (a_1,b_1) \frown (a_2,b_2) \cdots \frown (a_m,b_m) \tag{12-52}$$

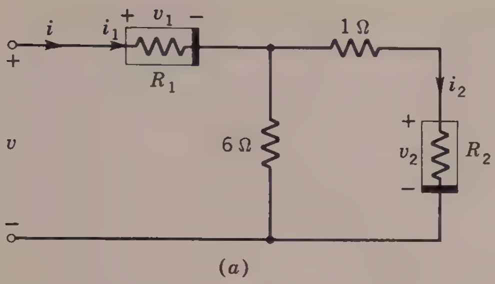
where the symbol $\frown$ denotes the standard intersection operation in mathematics.

In general, for some segment combinations Eq. (12-52) may not have a common interval, in which case we shall denote it by[1]

$$D(k_1,k_2, \ldots ,k_m) = 0 \tag{12-53}$$

Equation (12-53) means that the DP plot does not contain any portion of the straight line corresponding to this particular segment combination. For simplicity, we refer to this conclusion by saying that segment $(k_1,k_2, \ldots ,k_m)$ does not exist.

[1] In mathematics, the interval 0 is called the "empty set."



j	k	$_kr_j$	$_kE_j$	$_kD_j(v)$	$_kD_j(i)$
R_1	1	1.0	1.5	$(-\infty, 0.5)$	$(-\infty, -1)$
	2	0.5	1.0	$(0.5, \infty)$	$(-1, \infty)$
R_2	1	2.0	3.0	$(-\infty, 1)$	$(-\infty, -1)$
	2	1.0	2.0	$(1, \infty)$	$(-1, \infty)$

(b)

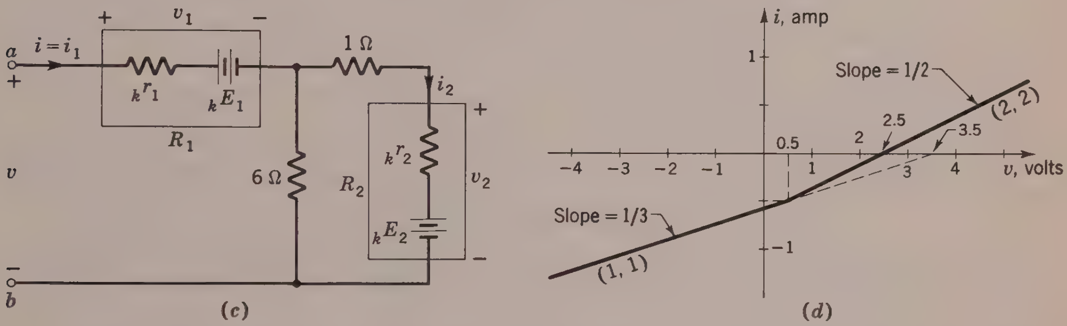


Fig. 12-9. The network for Example 1 and its DP plot as determined by the iterative piecewise-linear method.

The iterative piecewise-linear method consists of iterating Eq. (12-52) n times. Let us illustrate this by some numerical examples.

EXAMPLES

1. Consider the same network as that of Fig. 12-4 but with the battery replaced by a pair of driving-point terminals, as shown in Fig. 12-9*a*. The table defining the nonlinear resistors is reproduced in Fig. 12-9*b* for convenience. To find the DP plot, we draw the iterative linear network shown in Fig. 12-9*c*. From this we derive the following DP plot equation:[1]

$$i = \frac{1}{\Delta}\{(7 + {}_kr_2)v - [(7 + {}_kr_2){}_kE_1 + 6({}_kE_2)]\} \qquad (12\text{-}54)$$

where

$$\Delta \equiv (6 + {}_kr_1)(7 + {}_kr_2) - 36 \qquad (12\text{-}55)$$

Since $m = 2$, $s_1 = 2$, $s_2 = 2$, we have to iterate four times. However, instead of substituting immediately the parameters into these equations, it is advisable to determine first the interval of definition for each segment combination. This is because some segment combinations may not exist, in which case there is no need to substitute the corresponding set of parameters.

---

[1] Observe that Eqs. (12-54) and (12-55) have been derived earlier with $v = 2.5$ volts in Eqs. (12-12) and (12-14).

To determine the interval of definition, let us solve for i_1 and i_2 of the nonlinear resistors in terms of the driving-point voltage v†; thus

$$i_1 = \frac{1}{\Delta}\{(7 + {}_kr_2)v - [(7 + {}_kr_2)_kE_1 + 6({}_kE_2)]\} \qquad (12\text{-}56)$$

$$i_2 = \frac{1}{\Delta}\{6v - [6({}_kE_1) + (6 + {}_kr_1)_kE_2]\} \qquad (12\text{-}57)$$

where Δ is as defined in Eq. (12-55).

Observe that these equations correspond to Eq. (12-47), where $x_1 = i_1$, $x_2 = i_2$, and $x = v$. Let us start iterating as follows:

Segment (1,1): $\Delta = (6 + 1)(7 + 2) - 36 = 27$

$$i_1 = \tfrac{1}{27}\{(7 + 2)v - [(7 + 2)(1.5) + 6(3)]\} = \tfrac{1}{3}v - \tfrac{3.5}{3} \qquad (12\text{-}58)$$

$$i_2 = \tfrac{1}{27}\{6v - [6(1.5) + (6 + 1)(3)]\} = \tfrac{2}{9}v - \tfrac{10}{9} \qquad (12\text{-}59)$$

Since ${}_1D_1(i) = (-\infty,-1)$, ${}_1D_2(i) = (-\infty,-1)$, we obtain

$-\infty < \tfrac{1}{3}v - \tfrac{3.5}{3} < -1$ or $-\infty < v < 0.5$

$-\infty < \tfrac{2}{9}v - \tfrac{10}{9} < -1$ or $-\infty < v < 0.5$

The interval of definition for segment (1,1) is, therefore,

$D(1,1) = (-\infty,0.5) \cap (-\infty,0.5) = (-\infty,0.5)$

Segment (1,2): $\Delta = (6 + 1)(7 + 1) - 36 = 20$

$$i_1 = \tfrac{1}{20}\{(7 + 1)v - [(7 + 1)(1.5) + 6(2)]\} = \tfrac{2}{5}v - \tfrac{6}{5}$$

$$i_2 = \tfrac{1}{20}\{6v - [6(1.5) + (6 + 1)(2)]\} = \tfrac{6}{20}v - \tfrac{23}{20}$$

Since ${}_1D_1(i) = (-\infty,-1)$ and ${}_2D_2(i) = (-1,\infty)$, we obtain

$-\infty < \tfrac{2}{5}v - \tfrac{6}{5} < -1$ or $-\infty < v < 0.5$

$-1 < \tfrac{6}{20}v - \tfrac{23}{20} < \infty$ or $0.5 < v < \infty$

The interval of definition for segment (1,2) is, therefore,

$D(1,2) = (-\infty,0.5) \cap (0.5,\infty) = 0$

Segment (2,1): $\Delta = (6 + 0.5)(7 + 2) - 36 = 22.5$

$$i_1 = \tfrac{1}{22.5}\{(7 + 2)v - [(7 + 2)(1) + 6(3)]\} = \tfrac{2}{5}v - \tfrac{6}{5}$$

$$i_2 = \tfrac{1}{22.5}\{6v - [6(1) + (6 + 0.5)(3)]\} = \tfrac{4}{15}v - \tfrac{17}{15}$$

---

† Alternately, we can solve for v_1 and v_2 in terms of v; i_1 and i_2 in terms of i; v_1 and v_2 in terms of i; or any appropriate combination of these.

Since ${}_2D_1(i) = (-1,\infty)$, ${}_1D_2(i) = (-\infty,-1)$, we obtain

$$-1 < \tfrac{2}{5}v - \tfrac{6}{5} < \infty \qquad \text{or} \qquad 0.5 < v < \infty$$

$$-\infty < \tfrac{4}{15}v - \tfrac{17}{15} < -1 \qquad \text{or} \qquad -\infty < v < 0.5$$

The interval of definition for segment (2,1) is, therefore,

$$D(2,1) = (0.5,\infty) \frown (-\infty,0.5) = 0$$

Segment (2,2): $\Delta = (6 + 0.5)(7 + 1) - 36 = 16$

$$i_1 = \tfrac{1}{16}\{(7 + 1)v - [(7 + 1)(1) + 6(2)]\}$$
$$= \tfrac{1}{2}v - \tfrac{5}{4}$$

$$i_2 = \tfrac{1}{16}\{6v - [6(1) + (6 + 0.5)(2)]\}$$
$$= \tfrac{6}{16}v - \tfrac{19}{16}$$

Since ${}_2D_1(i) = (-1,\infty)$ and ${}_2D_2(i) = (-1,\infty)$, we obtain

$$-1 < \tfrac{1}{2}v - \tfrac{5}{4} < \infty \qquad \text{or} \qquad 0.5 < v < \infty$$

$$-1 < \tfrac{6}{16}v - \tfrac{19}{16} < \infty \qquad \text{or} \qquad 0.5 < v < \infty$$

The interval of definition for segment (2,2) is, therefore,

$$D(2,2) = (0.5,\infty) \frown (0.5,\infty) = (0.5,\infty)$$

The above results show that only two of the four segment combinations have a nonzero interval of definition. This means that the DP plot has only two segments, namely, segment (1,1) and segment (2,2). Substituting the parameters corresponding to segment (1,1) into Eq. (12-54), we obtain

$$i = \tfrac{1}{3}v - \tfrac{3.5}{3} \qquad -\infty < v < 0.5 \tag{12-60}$$

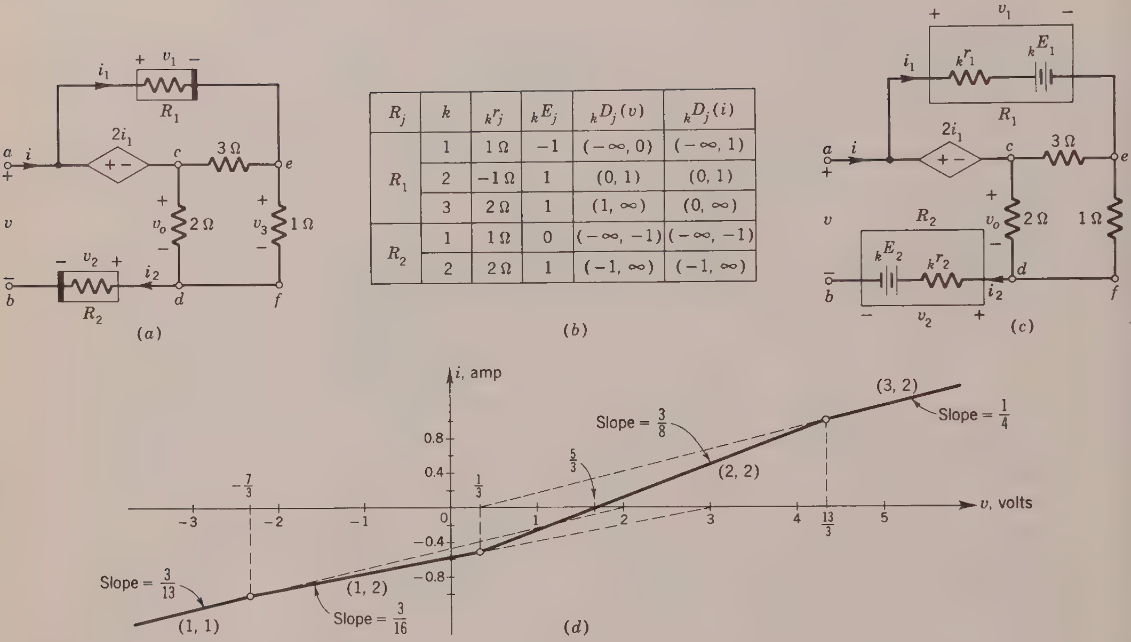
Substituting the parameters corresponding to segment (2,2) into Eq. (12-54), we obtain

$$i = \tfrac{1}{2}v - \tfrac{5}{4} \qquad 0.5 < v < \infty \tag{12-61}$$

From Eqs. (12-60) and (12-61) we obtain the DP plot shown in Fig. 12-9*d*. Observe that when $v = 2.5$, $i = 0$. This agrees with the operating point of the network shown in Fig. 12-4*a*.

At this point it is instructive to go back and read once more the general discussion of the iterative piecewise-linear method immediately preceding this example. Observe that Eq. (12-54) corresponds to Eq. (12-46). Similarly, Eqs. (12-56) and (12-57) correspond to Eq. (12-47). Observe also that the above iterative computations are actually equivalent to substituting the pertinent parameters into Eq. (12-49). For example, corresponding to Eqs. (12-48) and (12-49), we have

$${}_kI_j^- < i_j < {}_kI_j^+ \qquad j = 1, 2 \tag{12-62}$$



R_j	k	$_kr_j$	$_kE_j$	$_kD_j(v)$	$_kD_j(i)$
R_1	1	1 Ω	−1	$(-\infty, 0)$	$(-\infty, 1)$
	2	−1 Ω	1	$(0, 1)$	$(0, 1)$
	3	2 Ω	1	$(1, \infty)$	$(0, \infty)$
R_2	1	1 Ω	0	$(-\infty, -1)$	$(-\infty, -1)$
	2	2 Ω	1	$(-1, \infty)$	$(-1, \infty)$

Fig. 12-10. The network for Example 2 and its DP plot as determined by the iterative piecewise-linear method.

and

$$\frac{_kI_j^- - {_kK_j}}{_kH_j} < v < \frac{_kI_j^+ - {_kK_j}}{_kH_j} \qquad j = 1, 2 \tag{12-63}$$

In fact, we could have obtained the intervals of definition by substituting into Eq. (12-63) the parameters corresponding to each segment combination. For example, for segment (1,1) we identify from Eqs. (12-58) and (12-59) the coefficients $_1H_1 = \frac{1}{3}$, $_1K_1 = -{}^{3.5}\!/_3$, $_1H_2 = \frac{2}{9}$, and $_1K_2 = -{}^{10}\!/_9$. Substituting these coefficients with $_1D_1(i) = (-\infty, -1)$ and $_1D_2(i) = (-\infty, -1)$ into Eq. (12-63), we obtain

$$\frac{-\infty - (-{}^{3.5}\!/_3)}{\frac{1}{3}} < v < \frac{-1 - (-{}^{3.5}\!/_3)}{\frac{1}{3}} \qquad \text{or} \qquad -\infty < v < 0.5$$

$$\frac{-\infty - (-{}^{10}\!/_9)}{\frac{2}{9}} < v < \frac{-1 - (-{}^{10}\!/_9)}{\frac{2}{9}} \qquad \text{or} \qquad -\infty < v < 0.5$$

This result is, of course, identical with that which we obtained earlier. Actually, there is really not much saving in manual labor by making use of Eq. (12-63). We are doing it here primarily to verify the general theory. The real value of the general expression is in the writing of a computer program to carry out the iteration scheme automatically.

2. Consider next the same network as that of Fig. 12-5*a* but with the battery replaced by a pair of driving-point terminals as shown in Fig. 12-10*a*. The table defining the nonlinear resistors is reproduced in Fig. 12-10*b* for convenience. The associated iterative linear network is shown in Fig. 12-10*c*.

   We first solve for i in terms of v to obtain the iterative DP plot equation; thus

$$i = \left[\frac{1}{\alpha}\left(-\frac{1}{2} + {}_kr_1\right)\right]v + \left\{{}_kE_1 - \frac{1}{\alpha}\left(-\frac{1}{2} + {}_kr_1\right)\left[\left(\frac{4}{3} + {}_kr_2\right){}_kE_1 + {}_kE_2\right]\right\} \tag{12-64}$$

where

$$\alpha \equiv \tfrac{1}{3} + \tfrac{4}{3}({}_kr_1) - \tfrac{1}{2}({}_kr_2) + ({}_kr_1)({}_kr_2) \tag{12-65}$$

To determine the intervals of definition, we solve for i_1 and i_2 in terms of the driving-point voltage v; thus

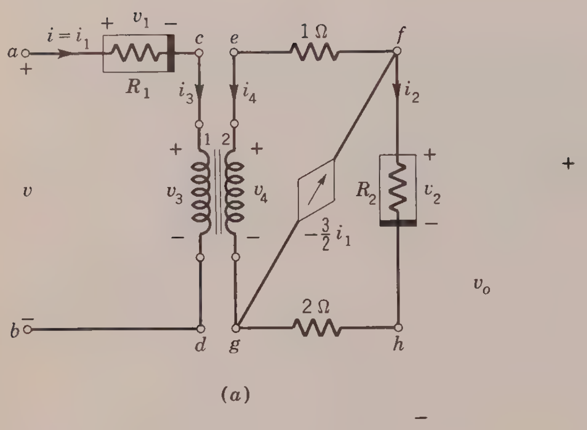
$$i_1 = \frac{1}{\alpha}v - \frac{1}{\alpha}\left[\frac{4}{3} + {}_kr_2){}_kE_1 + {}_kE_2\right] \tag{12-66}$$

$$i_2 = \frac{1}{\alpha}\left(-\frac{1}{2} + {}_kr_1\right)v + \left\{{}_kE_1 - \frac{1}{\alpha}\left(-\frac{1}{2} + {}_kr_1\right)\left[\left(\frac{4}{3} + {}_kr_2\right){}_kE_1 + {}_kE_2\right]\right\} \tag{12-67}$$

Equations (12-66) and (12-67) correspond to Eq. (12-47). Since $m = 2$, $s_1 = 3$, and $s_2 = 2$, we must iterate six times. The resulting intervals of definitions are tabulated in Table 12-3, together with the DP plot equations of the pertinent segments. The corresponding DP plot is shown in Fig. 12-10*d*.

TABLE 12-3 *Summary of results for defining the DP plot of Example 2.*

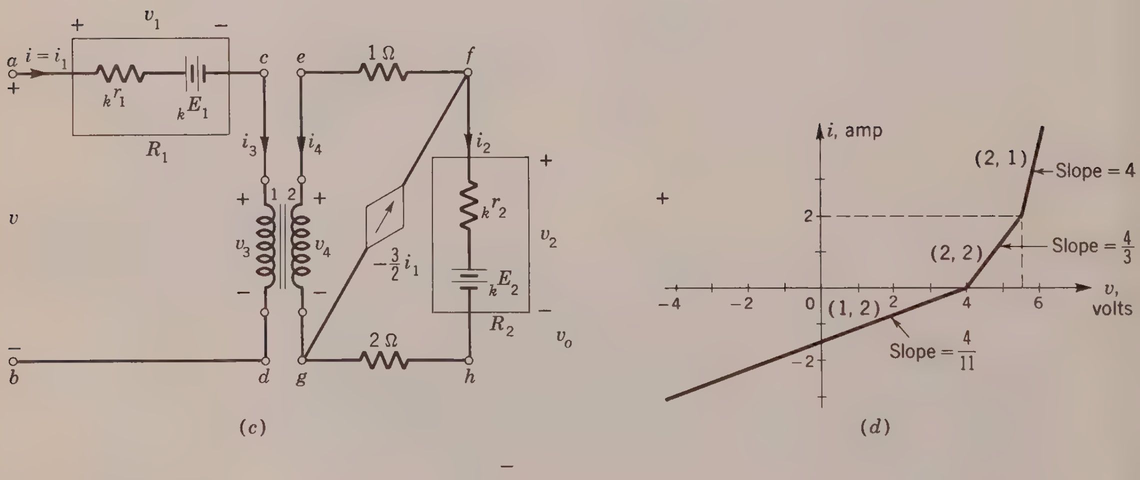
Segment	Interval of v for which $i_1 \in {}_kD_1(i)$ Eq. (12-66)	Interval of v for which $i_2 \in {}_kD_2(i)$ Eq. (12-67)	Common interval of definition $D(k_1,k_2)$	Equation of DP plot Eq. (12-64)
(1,1)	$(-\infty, -\tfrac{1}{6})$	$(-\infty, -\tfrac{7}{3})$	$(-\infty, -\tfrac{7}{3})$	$i = \tfrac{3}{13}v - \tfrac{6}{13}$
(2,1)	$(-\tfrac{1}{6}, \tfrac{7}{3})$	$(-\infty, -1)$	0	
(3,1)	$(\tfrac{7}{3}, \infty)$	$(-\infty, -\tfrac{11}{3})$	0	
(1,2)	$(-\infty, \tfrac{1}{3})$	$(-\tfrac{7}{3}, \infty)$	$(-\tfrac{7}{3}, \tfrac{1}{3})$	$i = \tfrac{3}{16}v - \tfrac{9}{16}$
(2,2)	$(\tfrac{1}{3}, \tfrac{13}{3})$	$(-1, \infty)$	$(\tfrac{1}{3}, \tfrac{13}{3})$	$i = \tfrac{3}{8}v - \tfrac{5}{8}$
(3,2)	$(\tfrac{13}{3}, \infty)$	$(-\tfrac{11}{3}, \infty)$	$(\tfrac{13}{3}, \infty)$	$i = \tfrac{1}{4}v - \tfrac{1}{12}$



(a)

R_j	k	${}_kr_j$	${}_kE_j$	${}_kD_j(v)$	${}_kD_j(i)$
R_1	1	4 Ω	3	(−∞, 3)	(−∞, 0)
	2	2 Ω	3	(3, ∞)	(0, ∞)
R_2	1	2 Ω	4	(−∞, 0)	(−∞, −2)
	2	1 Ω	2	(0, ∞)	(−2, ∞)

(b)



(c) (d)

Fig. 12-11. The network for Example 3 and its DP plot as determined by the iterative piecewise-linear method.

3. As a final example, consider the same network as that of Fig. 12-6*a* but with the battery replaced by a pair of driving-point terminals as in Fig. 12-11*a*. The table defining the nonlinear resistors is reproduced in Fig. 12-11*b* for convenience. The associated iterative linear network is shown in Fig. 12-11*c*.

   From the iterative linear network we derive the DP plot equation

$$i = \frac{1}{\alpha}\{4v - [4({}_kE_1) + 2({}_kE_2)]\} \tag{12-68}$$

where

$$\alpha = 4({}_kr_1) - 2({}_kr_2) - 3 \tag{12-69}$$

The equation for determining the intervals of definition are found to be

$$i_1 = \frac{4}{\alpha} v - \frac{1}{\alpha} [4({}_kE_1) + 2({}_kE_2)] \tag{12-70}$$

$$i_2 = -\frac{4}{\alpha} v + \frac{1}{\alpha} [4({}_kE_1) + 2({}_kE_2)] \tag{12-71}$$

Since $m = 2$, $s_1 = 2$, $s_2 = 2$, we have to iterate four times. The results are tabulated in Table 12-4, together with the pertinent DP plot equations. The resulting DP plot is shown in Fig. 12-11*d*.

Exercise 1: Show that the operating points of Examples 1, 2, and 3 of the preceding section can be obtained directly from the DP plots obtained in this section.

Exercise 2: Derive Eqs. (12-64) to (12-67) and verify the results in Table 12-3.

Exercise 3: Derive Eqs. (12-68) to (12-71) and verify the results in Table 12-4.

12-4 DETERMINATION OF TC PLOTS

The iterative piecewise-linear method presented in the preceding section can also be used to find the TC plot of any nonlinear network in the canonic form. As usual, we first draw the associated iterative linear network. However, instead of deriving the driving-point equation, we derive the specified transfer characteristics for the output variable y in terms of the driving-point variable x. This can always be written in the form

$$y = Tx + K \tag{12-72}$$

where y is either an output current or an output voltage and x is either the driving-point current or the driving-point voltage. The constants T and K are functions of ${}_kr_j$ and ${}_kE_j$. Corresponding to

TABLE 12-4 *Summary of results for defining the DP plot of Example 3.*

Segment	Interval of v for which $i_1 \in {}_kD_1(i)$ Eq. (12-70)	Interval of v for which $i_2 \in {}_kD_2(i)$ Eq. (12-71)	Common interval of definition $D(k_1,k_2)$	Equation of DP plot Eq. (12-68)
(1,1)	$(-\infty,5)$	$(\frac{19}{2},\infty)$	0	
(1,2)	$(-\infty,4)$	$(-\infty,\frac{19}{2})$	$(-\infty,4)$	$i = \frac{4}{11}v - \frac{16}{11}$
(2,1)	$(5,\infty)$	$(\frac{11}{2},\infty)$	$(\frac{11}{2},\infty)$	$i = 4v - 20$
(2,2)	$(4,\infty)$	$(-\infty,\frac{11}{2})$	$(4,\frac{11}{2})$	$i = \frac{4}{3}v - \frac{16}{3}$

each set of parameter values associated with a segment combination, Eq. (12-72) represents a straight line in the transfer plane y versus x. Since there are n combinations, where n is defined by Eq. (12-1), there are, accordingly, n straight lines, and the TC plot is made up of portions of these straight lines. Just as for DP plots the crucial step here is to determine the interval of definition for each segment of the TC plot.

Since the TC plot is simply the locus of operating points of the desired output variable plotted as a function of the driving-point variable i or v, the intervals of definition can be determined by exactly the same method discussed in the preceding section for DP plots.

EXAMPLES

1. Consider the network shown in Fig. 12-9 again, but this time let us determine the TC plot between v_2 and v. Instead of deriving the driving-point equation between i and v, we solve for v_2 versus v; thus

$$v_2 = \left[\frac{1}{\Delta}\,6({}_kr_2)\right]v - \frac{1}{\Delta}\,({}_kr_2)[6({}_kE_1) + (6 + {}_kr_1){}_kE_2] + {}_kE_2 \tag{12-73}$$

where Δ is defined earlier in Eq. (12-55).

To find the interval of definition, we must derive the relationships i_1 versus v and i_2 versus v. Since this involves the same procedure used in determining the intervals of definition of the DP plots, Eqs. (12-56) and (12-57) are applicable here, and the same intervals of definition will result, namely,

Segment (1,1): $D(1,1) = (-\infty, 0.5)$

Substituting the parameters of segment (1,1) into Eq. (12-73), we obtain

$$v_2 = \tfrac{4}{9}v + \tfrac{7}{9}$$

Segment (2,2): $D(2,2) = (0.5, \infty)$

Substituting the parameters of segment (2,2) into Eq. (12-73), we obtain

$$v_2 = \tfrac{3}{8}v + \tfrac{13}{16}$$

The resulting TC plot is shown in Fig. 12-12.

2. Consider the network shown in Fig. 12-10*a* again, but this time we want the TC plot v_o versus v. The equation relating v_o

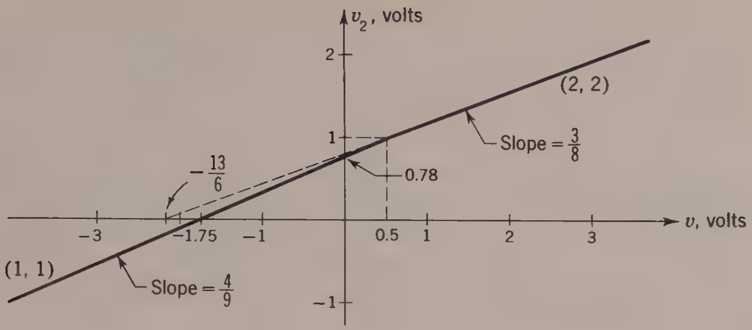


Fig. 12-12. The TC plot for Example 1.

versus v is derived from the iterative linear network of Fig. 12-10*c*.

$$v_o = \frac{1}{3\alpha}[-5 + 4({}_kr_1)]v + \frac{1}{3\alpha}\{[3({}_kr_2) + 8]_kE_1 + [5 - 4({}_kr_1)]_kE_2\} \qquad (12\text{-}74)$$

where α is as defined in Eq. (12-65).

Again, the intervals of definition for this TC plot have already been found in Sec. 12-3 as tabulated in Table 12-3.

Segment (1,1): $D(1,1) = (-\infty, -\frac{7}{3})$

Substituting the parameters for segment (1,1) into Eq. (12-74), we obtain

$$v_o = -\tfrac{2}{13}v - \tfrac{22}{13}$$

Segment (1,2): $D(1,2) = (-\frac{7}{3}, \frac{1}{3})$

Substituting the parameters for segment (1,2) into Eq. (12-74), we obtain

$$v_o = -\tfrac{1}{8}v - \tfrac{13}{8}$$

Segment (2,2): $D(2,2) = (\frac{1}{3}, \frac{13}{3})$

Substituting the parameters for segment (2,2) into Eq. (12-74), we obtain

$$v_o = \tfrac{3}{4}v - \tfrac{23}{12}$$

Segment (3,2): $D(3,2) = (\frac{13}{3}, \infty)$

Substituting the parameters for segment (3,2) into Eq. (12-74), we obtain

$$v_o = \tfrac{1}{6}v + \tfrac{11}{18}$$

The resulting TC plot is shown in Fig. 12-13.

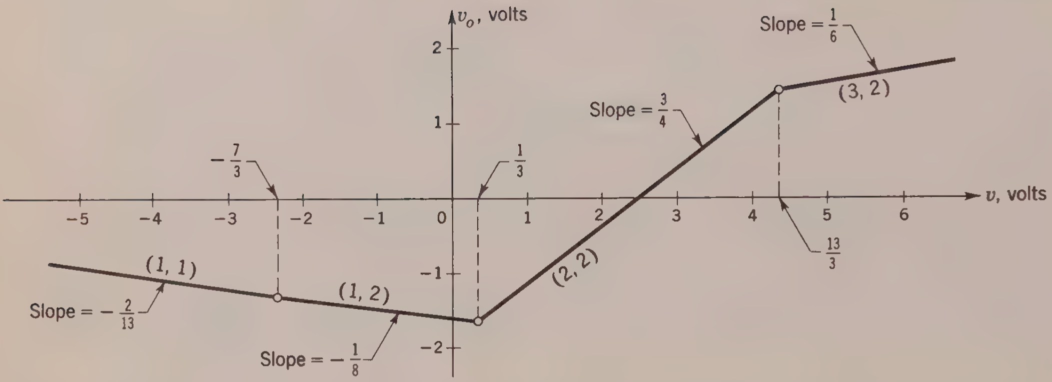


Fig. 12-13. The TC plot for Example 2.

3. As a final example, consider the network shown in Fig. 12-11 again, but this time we want to find the TC plot v_o versus v. We can derive this relationship from the iterative linear network

$$v_o = \left[-\frac{4}{\alpha}(2 + {}_kr_2)\right]v + \frac{1}{\alpha}\{(2 + {}_kr_2)[4({}_kE_1) + 2({}_kE_2)]\} + {}_kE_2 \tag{12-75}$$

where α is as defined in Eq. (12-69).

The intervals of definition for this TC plot have already been found in Sec. 12-3 as tabulated in Table 12-4. Thus we have

Segment (1,2): $D(1,2) = (-\infty,4)$

Substituting the parameters for segment (1,2) into Eq. (12-75), we obtain

$v_o = -{}^{12}\!/_{11}v + {}^{70}\!/_{11}$

Segment (2,1): $D(2,1) = ({}^{11}\!/_{2},\infty)$

Substituting the parameters for segment (2,1) into Eq. (12-75), we obtain

$v_o = -16v + 84$

Segment (2,2): $D(2,2) = (4,{}^{11}\!/_{2})$

Substituting the parameters for segment (2,2) into Eq. (12-75), we obtain

$v_o = -4v + 18$

The resulting TC plot is shown in Fig. 12-14.

From the above examples, it is clear that the only difference between the iterative piecewise-linear method for determining DP plots and TC plots is in the choice of the output variable. The output variable for DP plots is by definition located at the driving-point terminals, whereas the output variable for TC plots can be located anywhere. In fact, we can actually consider DP plots as special cases of TC plots. Observe also that the crucial step involving the determination of intervals of definition is independent of whether we are solving for a DP plot or a TC plot. As a consequence of this, we can conclude that *the DP plot and the TC plot of any network (as a function of the same driving-point variable) must have the same number of segments.*

It is important to realize that the iterative piecewise-linear method can be implemented in two parts. The first part involves a solution of network equations for the iterative linear network. This can be obtained by any convenient linear technique. The second part involves the iteration and comparison procedures over all possible segment combinations. The first part is conceptual, while the second part is mechanical. It is obvious, therefore, that the second part could be easily implemented by a computer, thereby avoiding the tedious part of the method.

Exercise 1: Derive Eqs. (12-73) to (12-75).

Exercise 2: Show that the procedure for deriving DP plots can be considered as a special case of the procedure for deriving TC plots. Give the flow chart and algorithm for implementing the iteration part of the method by a computer.

Exercise 3: Show why the iterative piecewise-linear method is valid for determining the operating point, DP plot, and TC plot of *any* nonlinear resistive network in the canonic form.

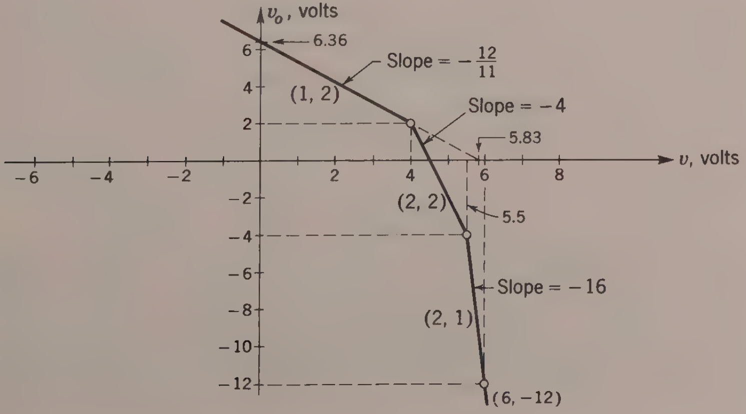


Fig. 12-14. The TC plot for Example 3.

Exercise 4: Show that in order to avoid using different equivalent circuits due to the presence of nonlinear resistors with vertical and horizontal segments, it is possible to replace these resistors by equivalent resistors having neither vertical nor horizontal segments.

Exercise 5: Show that the solution of any ac-resistive network containing only one ac voltage source or one ac current source can be easily found by the iterative piecewise-linear method. NOTE: The network may contain any number of dc independent sources and controlled sources.

12-5 COMPUTER ANALYSIS OF RESISTIVE NONLINEAR NETWORKS

The iterative piecewise-linear method is the most general method currently available for analyzing arbitrary resistive nonlinear networks in the canonic form. This method has three significant advantages over the numerical techniques described earlier in Chap. 4. First, whereas most numerical techniques are applicable only to networks with unique solutions, the present method is applicable to multivalued networks. Second, it is often impractical to derive the nonlinear equations describing the network before numerical analysis can be undertaken; however, it is always possible to derive the equations for the iterative piecewise-linear method. Third, a numerical technique may or may not converge, but the iterative piecewise-linear method will always give all solutions at the end of the iteration, even though it may take a great many iterations for complex networks. Perhaps the only disadvantage of the present method is that the computer time could be excessive for a network containing many nonlinear resistors. It is hopeful, however, that this disadvantage will soon be overcome by more efficient bookkeeping techniques for eliminating a large number of inconsistent segment combinations.

Two general user-oriented computer programs for implementing the iterative piecewise-linear method have been developed and can be obtained from the author. Since no programming background is assumed of the reader, we shall describe these two programs only from the user's point of view. The details of the programs are available in separate reports.

12-5-1 SEMIAUTOMATIC GENERAL ANALYSIS PROGRAM[1]

In this approach the user must first perform a linear network analysis and derive all pertinent equations needed for determining the operating point, the DP plot, and the TC plot. This corresponds to the first part of the iterative piecewise-linear method. The second part of the method is taken over by a general analysis

---

[1] L. O. Chua and W. H. Stellhorn, Computer-aided Analysis and Synthesis of Multivalued Memoryless Networks, *Purdue University Report* TR-EE-67-8, June, 1967.

program which will automatically perform all iterations needed. The *input* to this program consists of (1) the equations derived from part 1 and (2) the slope, intercept, and interval of definition of each segment of the nonlinear resistors. The *output* of the program will give the operating points and the DP plot or the TC plot, whichever is called for. This method is semiautomatic because the user must do part of the work. For complex networks even this step could be quite time-consuming. It has the advantage, however, that once the pertinent equations have been derived, we can change the v-i curves of the nonlinear resistors and ask for the solution without rederiving any equation. Hence, this approach is particularly well-suited where a detailed qualitative and quantitative analysis of networks having identical topology, such as a bridge network, is desired. In fact, because the equations are functions of the parameters defining the nonlinear resistors, the effect of parameter variations on the network's performance can be determined. Because the program is not completely automatic, it can be kept to a small size, and hence does not require a very large computer. Moreover, the computation time is relatively short.

12-5-2 COMPLETELY AUTOMATIC MECA PROGRAM[1]

In this approach the user does not have to derive a single equation. Only the network connection diagram and the element characteristics need be specified. The basic technique consists of solving a linear network with *numerical parameters* many times. In other words, instead of assuming symbolic parameters such as ${}_kr_j$ and ${}_kE_j$, the numerical values of the parameter for each segment combination are assumed from the outset, leaving a linear network whose solutions can be easily obtained. Since the computer is very fast, for circuits of moderate complexity, we can afford to repeat the analysis as many times as there are segment combinations.

The computer program currently available for the completely automatic case is called Meca. It is an acronym for multivalued electronic circuit analysis program.

As an example of the use of this program, consider the network shown in Fig. 12-15*a*. This circuit is called the Schmitt trigger and is one of the most commonly used nonlinear circuits in practice. The usefulness of this circuit can be seen by deriving the v_o-vs.-v_i TC plot from the collector of transistor T_2 to the base of transistor T_1. For most practical purposes the two transistors are usually identical.

---

[1] L. O. Chua and P. A. Medlock, MECA—A User-Oriented Computer Program for Analyzing Resistive Nonlinear Networks, Volume 1: User's Manual, *Purdue University Report* TR-EE-69-7, April, 1969.

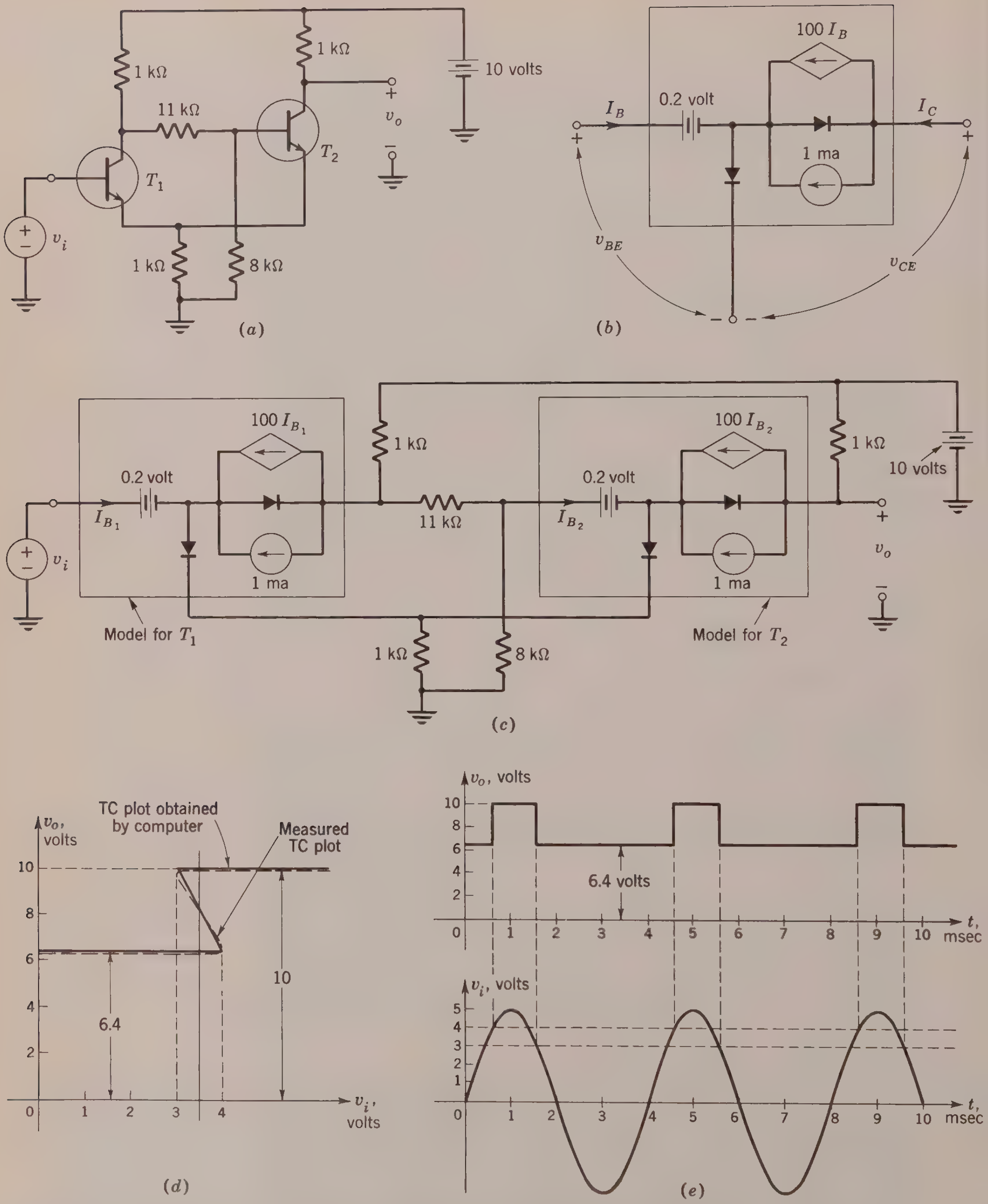


Fig. 12-15. The TC plot of the Schmitt trigger can be obtained automatically by using the Meca program.

In order to use the Meca program, we must first recast the Schmitt trigger circuit into the canonic form by substituting a network model for each of the two identical transistors. For this example, we have chosen the *n-p-n* transistor model C_E from Chap. 11. The two parameters needed in this model were obtained directly from the characteristic curves of the transistors used in the actual circuit. The resulting model is shown in Fig. 12-15*b* with $V_{BEO} = 0.2$ volt and $I_{CEO} = 1$ ma. Substituting this model in place of the two transistors, we obtain the Schmitt trigger circuit in the canonic form as shown in Fig. 12-15*c*. Once this is done, the next step is to label all nodes and elements with numbers in any arbitrary order. These labels are used only for describing the connection diagram to the computer. For example, the position of a nonlinear resistor R_j is simply specified by the label of the nodes across which R_j is connected. The location of a current-controlled current source is specified by the label of the nodes across which the controlled source is connected and the label of the element whose current is the controlling quantity. In other words, only the minimum amount of information needed to specify the circuit unambiguously is required by the Meca program. We may liken this to composing the shortest possible telegram for describing a network connection diagram so that the receiver may be able to reconstruct the network without ambiguity. With this minimum amount of information as the input, the v_o-vs.-v_i TC plot is automatically computed by the Meca program, and the result is printed out in the form of a piecewise-linear curve, such as that shown by the solid curve in Fig. 12-15*d*. As a comparison, the actually measured TC plot is shown by the dotted curve in the same figure. Notice that except in the vicinity of the two breakpoints, the two TC plots are identical for most practical purposes. This observation should convince the reader of the validity of the models chosen.

From this TC plot we observe that the collector output voltage $v_o(t)$ will jump from 6.4 to 10 volts whenever the input voltage $v_i(t)$ increases beyond 4 volts. When this voltage is then decreased, the output voltage remains at 10 volts until $v_i(t)$ decreases below 3 volts, at which time the output drops back to 6.4 volts. Since the forward and reverse critical voltages, called the trigger voltages, are not equal to each other, this circuit is said to possess a *hysteresis property*. An example of this property is clearly shown by the pair of input-output waveforms shown in Fig. 12-15*e*. Notice that by controlling the length of the hysteresis gap, a nonsymmetrical square wave can be easily generated. In fact, one common application of the Schmitt trigger is converting a sinusoidal wave into a square wave. The Schmitt trigger circuit has many other applications in the field of instrumentation and control. Most of these applications are based on the two-distinct-level property of the TC plot, which makes it a very good threshold detector.

12-6 TC PLOT SYNTHESIS BY THE PIECEWISE-LINEAR METHOD

Several graphical methods were presented in Chap. 8 for realizing a prescribed TC plot. We shall now show that some of these methods can be more easily implemented using the piecewise-linear method. Moreover, the effect of the element parameters on the resulting TC plot can usually be derived in exact analytic form. Another advantage of the piecewise-linear synthesis method is that since specific formulas can be derived, it is more amenable to a digital computer solution.

12-6-1 TC PLOT SYNTHESIS BY NONLINEAR VOLTAGE DIVIDERS

Any prescribed TC plot can be synthesized by the two nonlinear voltage-divider networks shown earlier in Fig. 8-15. It suffices for us to consider the case where R_2 is the nonlinear resistor. Replacing R_2 by its iterative Thévenin equivalent circuit, we obtain the circuit shown in Fig. 12-16*a*. As usual, the value of the linear resistor R_1 can be arbitrarily chosen and can therefore be assumed as given. Our problem is to specify the piecewise-linear v_2-i_2 curve for R_2. An examination of the circuit in Fig. 12-16*a* shows that since there is only one nonlinear resistor, the TC plot of this network must have exactly the same number of segments as that of R_2. This follows because each segment combination of the iterative piecewise-linear method consists of only one segment, and hence there is always a nonempty common interval of definition. Furthermore, since $v_o = v_2$ in this circuit, the coordinate of v_2 at each breakpoint of the v_2-i_2 curve for R_2 must be identical with the coordinate of the output voltage v_o of the corresponding breakpoint of the prescribed TC plot. Hence, given an arbitrary TC plot as that shown in Fig. 12-16*b*, we can immediately determine the interval of definition for each segment of the v_2-i_2 curve by aligning the i_2 axis with the v_i axis of the TC plot as shown in Fig. 12-16*c*. It remains for us, therefore, to find the value of ${}_kr_2$ and ${}_kE_2$ for each segment of the v_2-i_2 curve over each interval of definition.

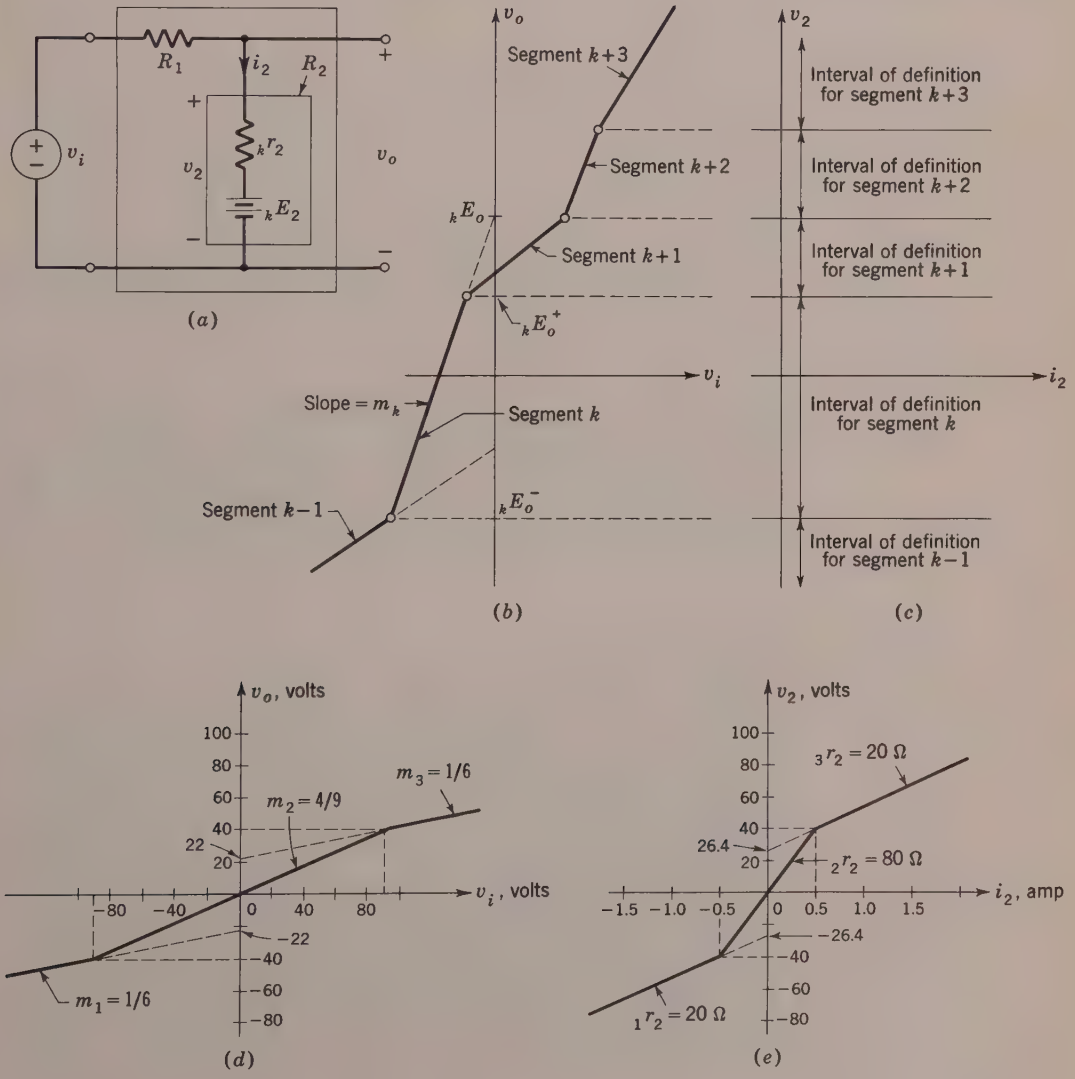
The equation describing each segment k of the TC plot of the voltage-divider network in Fig. 12-16*a* is easily derived as

$$v_o = \frac{{}_kr_2}{R_1 + {}_kr_2} v_i + \frac{R_1}{R_1 + {}_kr_2} {}_kE_2 \tag{12-76}$$

Denoting the slope and intercept of segment k of the prescribed TC plot by m_k and ${}_kE_o$, we obtain the equation of segment k as follows:

$$v_o = m_k v_i + {}_kE_o \tag{12-77}$$

Fig. 12-16. The synthesis of a nonlinear voltage-divider network with a prescribed TC plot by the piecewise-linear method.



Since Eqs. (12-76) and (12-77) describe the same TC plot, the respective coefficients must be equal to each other, that is,

$$\frac{{}_k r_2}{R_1 + {}_k r_2} = m_k \tag{12-78}$$

$$\frac{R_1({}_k E_2)}{R_1 + {}_k r_2} = {}_k E_o \tag{12-79}$$

Solving for ${}_k r_2$ and ${}_k E_2$ from Eqs. (12-78) and (12-79), we obtain

$${}_k r_2 = \frac{R_1 m_k}{1 - m_k} \tag{12-80}$$

$${}_k E_2 = \frac{{}_k E_o}{1 - m_k} \tag{12-81}$$

Equations (12-80) and (12-81) specify the values of ${}_k r_2$ and ${}_k E_2$ for each segment k of the v_2-i_2 curve directly in terms of R_1, m_k, and ${}_k E_o$. These, together with the intervals of definitions obtained earlier, completely specify the required v_2-i_2 curve. From Eq. (12-80) we find that ${}_k r_2 \geq 0$ if, and only if,

$$0 \leq m_k \leq 1 \tag{12-82}$$

provided $R_1 > 0$. This confirms our earlier observation in Sec. 8-4-1 that the v_2-v_2 curve for R_2 is monotonically increasing if, and only if, the slope of each segment of the prescribed TC plot is positive and less than unity. If Eq. (12-82) is not satisfied, the v_2-i_2 curve will either be nonmonotonic or multivalued.

EXAMPLE

Suppose the prescribed TC plot is given by Fig. 12-16*d*. The pertinent parameters are segment 1, $m_1 = \frac{1}{6}$, ${}_1 E_o = -22$ volts; segment 2, $m_2 = \frac{4}{9}$, ${}_2 E_o = 0$ volt; and segment 3, $m_3 = \frac{1}{6}$, ${}_3 E_o = 22$ volts. To determine the v_2-i_2 curve of R_2 in order to realize this TC plot, let us choose (arbitrarily) $R_1 = 100\ \Omega$ and substitute these parameters into Eqs. (12-80) and (12-81). The corresponding results are given by segment 1, ${}_1 r_2 = 20\ \Omega$, ${}_1 E_2 = -26.4$ volts; segment 2, ${}_2 r_2 = 80\ \Omega$, ${}_2 E_2 = 0$ volt; and segment 3, ${}_3 r_2 = 20\ \Omega$, ${}_3 E_2 = 26.4$ volts. The v_2-i_2 curve corresponding to these parameters is shown in Fig. 12-16*e*. Notice that the breakpoint voltages of both the v_i-v_o TC plot and the corresponding i_2-v_2 curve are identical.

12-6-2 JOINTLY PRESCRIBED DP PLOT, TC PLOT, AND LOAD v-i CURVE SYNTHESIS

We are now in a position to derive the lattice network synthesis theorem which was presented earlier in Sec. 8-5 without proof. In order to apply the iterative piecewise-linear theory, the prescribed DP plot, TC plot, and load v_L-i_L curve are assumed to be approximated by piecewise-linear segments. Hence, each segment j of the $v_o = T(v)$ TC plot can be described by

$$v_o = m_j v + E_j \qquad \alpha_j < v < \beta_j \qquad j = 1, 2, \ldots, n \tag{12-83}$$

where m_j and E_j are the slope and voltage intercept of segment j, and (α_j,β_j) is the corresponding interval of definition. Similarly, each segment j of the $i = g(v)$ DP plot can be described by

$$i = g_j v + I_j \qquad \alpha'_j < v < \beta'_j \qquad j = 1, 2, \ldots, n' \tag{12-84}$$

where g_j and I_j are the slope and intercept of segment j and (α'_j,β'_j) is its interval of definition. Notice that both Eqs. (12-83) and (12-84) involve the same independent variable v. As will be obvious soon, it is desirable that the third specification, namely, the load characteristic $i_L = g_L(v_L)$, also be expressed in terms of this common variable v. This can be easily done because $v_L = v_o = T(v)$, and hence, this specification can be recast into an *equivalent load specification* $i_L = g_L(T(v))$ by a simple graphical composition between the load v_L-i_L curve and the TC plot. Let this composition be done, and let each segment j of the resulting piecewise-linear curve be described by

$$i_L = ({}_jK_L)v + {}_jI_L \qquad \alpha''_j < v < \beta''_j \qquad j = 1, 2, \ldots, n'' \tag{12-85}$$

where ${}_jK_L$ and ${}_jI_L$ are the slope and intercept of segment j of this curve and (α''_j,β''_j) is its interval of definition.

Since Eqs. (12-83) to (12-85) are obtained from the specifications, they can be considered given at the outset as an equivalent form of specification. Our problem is to derive the design equations for finding the resistors R_A and R_B of the symmetric lattice network given in Fig. 8-17*a*. Replacing these resistors by their iterative Thévenin equivalent circuit, we obtain the circuit shown in Fig. 12-17. Observe that since the two resistors in the series arms are identical, they are described by the same parameters ${}_jr_1$ and ${}_jE_1$. Similarly, the two resistors in the shunt arms are described by ${}_jr_2$ and ${}_jE_2$. Our problem consists of two parts: (1) we

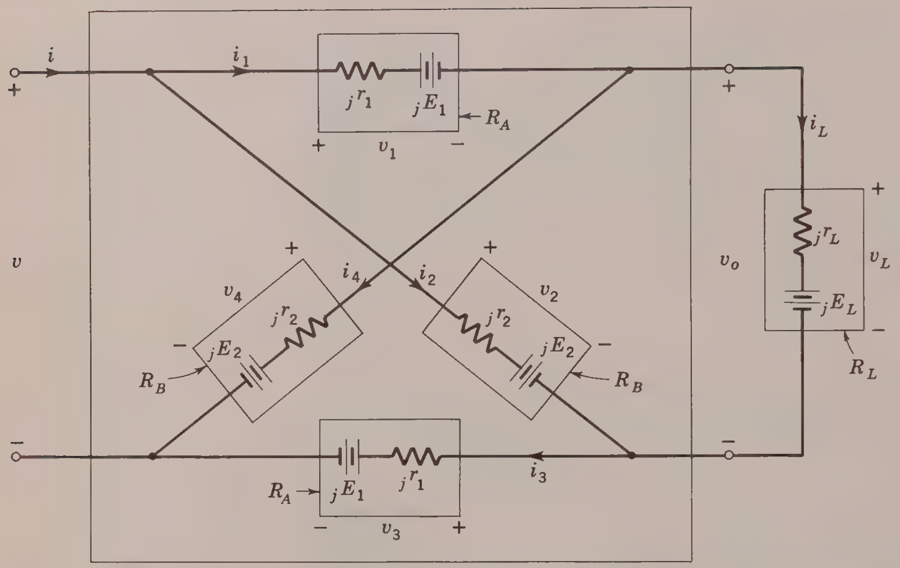


Fig. 12-17. The iterative piecewise-linear equivalent circuit associated with the symmetric nonlinear lattice network.

must determine the number of segments of R_A and R_B and the interval of definition for each of these segments, and (2) we must determine the values of $_jr_1$, $_jE_1$ for R_A and $_jr_2$, $_jE_2$ for R_B.

To determine the number of segments of R_A and R_B, let us consider the special case first where $n = n' = n''$, and where the intervals of definition (α_j,β_j), (α'_j,β'_j), and (α''_j,β''_j) are identical for each segment j; namely,

$$\alpha_j = \alpha'_j = \alpha''_j = a_j \tag{12-86}$$

$$\beta_j = \beta'_j = \beta''_j = b_j \tag{12-87}$$

Under these conditions, the three piecewise-linear curves described by Eqs. (12-83) to (12-85) will have the same set of breakpoints. This means that at each value of $v = \alpha_j = \alpha'_j = \alpha''_j$ or $v = \beta_j = \beta'_j = \beta''_j$, all three curves will be operating simultaneously at a breakpoint. This behavior is possible only if all resistors R_A and R_B are also simultaneously operating at a breakpoint. We therefore conclude that if Eqs. (12-86) and (12-87) are satisfied, then R_A and R_B must have the same number of segments $n = n' = n''$.

In practice, the intervals of definitions (α_j,β_j), (α'_j,β'_j), and (α''_j,β''_j) are generally different, and hence Eqs. (12-86) and (12-87) are generally not satisfied. However, by subdividing some segments into two or more subsegments, it is always possible to satisfy these equations. All we need to do is to align the vertical axis of the TC plot, the DP plot, and the equivalent load characteristic along their vertical axis and draw a guideline through *each* breakpoint of *each* of these three curves. If we let the distance between each pair of adjacent vertical lines be the interval of definition of a subsegment, then the entire v axis will be divided into N subintervals, where

$$\max(n,n',n'') \le N \le (n + n' + n'') \tag{12-88}$$

In view of the above procedure, we can be sure that Eqs. (12-86) and (12-87) will always be satisfied, and hence, resistors R_A and R_B must have exactly N segments. For example, consider the same specifications given earlier in Fig. 8-17*b* to *d*, which we redraw in Fig. 12-18*a* to *c* for convenience. Our first step is to perform a graphical composition between the first two curves to obtain the equivalent load specification shown in Fig. 12-18*d*. Observe that whereas the original curves have three segments each, the equivalent load specification has five segments. Hence, $n = n' = 3$, and $n'' = 5$. Observe also that the intervals of definition for each of these curves are different. However, if we draw a vertical guideline through each and every breakpoint of the TC plot, DP plot, and the equivalent load characteristic, as shown in Fig. 12-18*b* to *d*, we would subdivide the v axis into six intervals, namely, $(-\infty,-5)$, $(-5,-3)$, $(-3,-1)$, $(-1,2)$, $(2,4\tfrac{2}{3})$, and $(4\tfrac{2}{3},\infty)$. If we introduce a *breakpoint* at the intersection between each segment and these vertical guidelines, then each of the curves in Fig. 12-18*b* to *d* can be interpreted as made up of six segments, even though some segments may fall on the same straight line. The reason for doing this is obviously to satisfy Eqs. (12-86) and (12-87). For future reference, the slope and intercept of each of the six-segment TC plot and DP plot of Fig. 12-18*b* and *c* are tabulated in columns 3 to 6 of Table 12-5. Corresponding to the six

TABLE 12-5 *Summary of pertinent parameters from the combined specifications.*

Segment number j	Range of driving-point voltage V	TC plot		DP Plot		Load v_L-i_L Curve	
		m_j	E_j	g_j	I_j	${}_jr_L$	${}_jE_L$
1	$(-\infty,-5)$	$\frac{1}{2}$	$\frac{3}{2}$	$\frac{2}{3}$	1	4	-1
2	$(-5,-3)$	$\frac{1}{2}$	$\frac{3}{2}$	$\frac{2}{3}$	1	1	-1
3	$(-3,-1)$	$\frac{1}{2}$	$\frac{3}{2}$	1	2	1	-1
4	$(-1,2)$	$-\frac{2}{3}$	$\frac{1}{3}$	1	2	1	-1
5	$(2,4\tfrac{2}{3})$	$\frac{3}{4}$	$-\frac{5}{2}$	1	2	1	-1
6	$(4\tfrac{2}{3},\infty)$	$\frac{3}{4}$	$-\frac{5}{2}$	$\frac{1}{2}$	$\frac{13}{3}$	2	-3

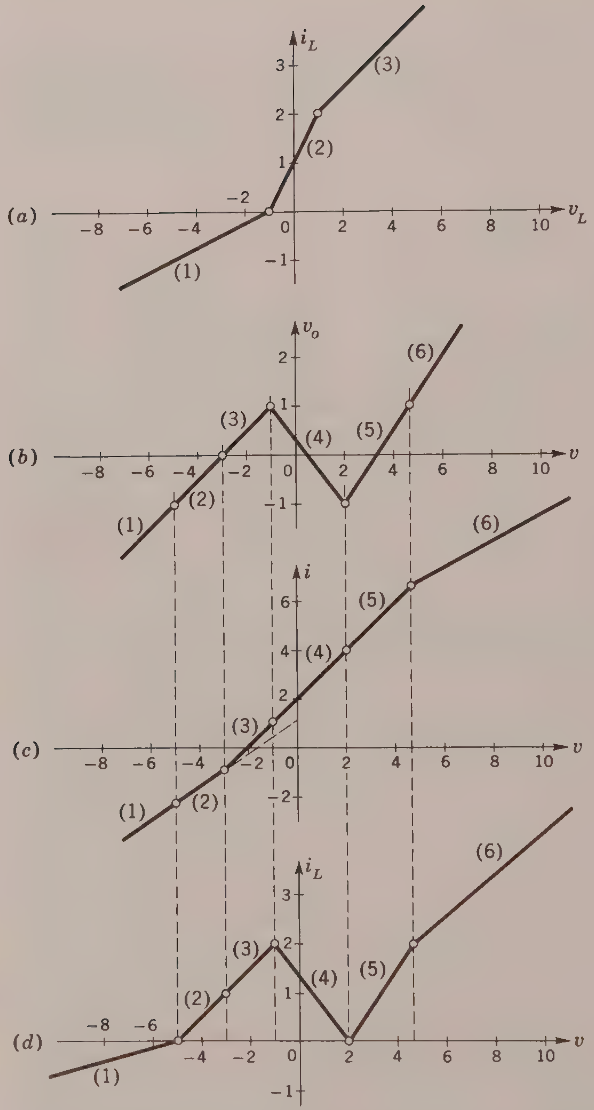


Fig. 12-18. An example of a jointly prescribed load characteristic, TC plot, DP plot, and the associated equivalent load specification.

intervals of driving-point voltage v, the original load resistor v_L-i_L curve in Fig. 12-18*a* can also be broken up into six intervals, but in terms of i_L. This is most simply done by superimposing the v_L-i_L curve of Fig. 12-18*a* on top of the equivalent load specification of Fig. 12-18*d*. Since the ordinates of these two curves are identical, we can identify the interval of voltage v_L of the original curve corresponding to the six intervals obtained earlier. For example, segment 1 of Fig. 12-18*d* is operating at $-\infty < i_L < 0$. The corresponding interval in Fig. 12-18*a* shows that when $-\infty < i_L < 0$, $-\infty < v_L < -1$, and hence the load resistor is operating at segment 1 during this interval. Similarly, corresponding to segments 2 to 5 in Fig. 12-18*d*, the load resistor is operating in segment 2. Finally, corresponding to segment 6 ($2 < i_L < \infty$) in Fig. 12-18*d*, the load resistor v_L-i_L curve in Fig. 12-18*a* is seen to be operating in segment 3. The values of ${}_jr_L$ and ${}_jE_L$ of the load v_L-i_L curve corresponding to the six intervals of definition are tabulated in columns 7 and 8 of Table 12-5.

As a summary, we found that the crucial step in this method consists of taking the graphical composition between the load v_L-i_L curve and the TC plot to obtain equivalent load specification. The three curves (TC plot, DP plot, and equivalent load specification) have the same independent variable v and can, therefore, be broken up into N common intervals (a_j,b_j), $j = 1, 2, \ldots, N$. During each interval, we are guaranteed that each resistor will be operating in one, and only one, segment. It makes sense, therefore, to compare each segment of the TC plot and DP plot of the iterative linear lattice network in Fig. 12-17 with the given specification in the same manner as that of the voltage divider.

By means of the symmetric lattice network theorem (Sec. 7-7-1), we obtain $v_1 = v_3$, $i_1 = i_3$, $v_2 = v_4$, and $i_2 = i_4$. Using these relationships, we derive the following equations from the network of Fig. 12-17:

$$v_o = \frac{({}_jr_2 - {}_jr_1){}_jr_L}{\Delta}v + \frac{2{}_jr_L[({}_jr_1)({}_jE_2) - ({}_jr_2)({}_jE_1)] + 2({}_jr_1)({}_jr_2)({}_jE_L)}{\Delta} \tag{12-89}$$

$$i = \frac{[{}_jr_1 + {}_jr_2 + 2({}_jr_L)]}{\Delta}v + \frac{({}_jr_1 - {}_jr_2){}_jE_L - 2({}_jr_1 + {}_jr_L){}_jE_2 - 2({}_jr_2 + {}_jr_L){}_jE_1}{\Delta} \tag{12-90}$$

where

$$\Delta \equiv ({}_jr_1 + {}_jr_2){}_jr_L + 2({}_jr_1)({}_jr_2) \tag{12-91}$$

If we equate the coefficients in Eqs. (12-89) and (12-90) with the corresponding coefficients in Eqs. (12-83) and (12-84), we obtain

$$\frac{({}_jr_2 - {}_jr_1){}_jr_L}{({}_jr_1 + {}_jr_2){}_jr_L + 2({}_jr_1)({}_jr_2)} = m_j \tag{12-92}$$

$$\frac{[{}_jr_1 + {}_jr_2 + 2({}_jr_L)]}{({}_jr_1 + {}_jr_2){}_jr_L + 2({}_jr_1)({}_jr_2)} = g_j \tag{12-93}$$

$$\frac{2({}_jr_L)[({}_jr_1)({}_jE_2) - ({}_jr_2)({}_jE_1)] + 2({}_jr_1)({}_jr_2)({}_jE_L)}{({}_jr_1 + {}_jr_2){}_jr_L + 2({}_jr_1)({}_jr_2)} = E_j \tag{12-94}$$

$$\frac{({}_jr_1 - {}_jr_2){}_jE_L - 2({}_jr_1 + {}_jr_L){}_jE_2 - 2({}_jr_2 + {}_jr_L){}_jE_1}{({}_jr_1 + {}_jr_2){}_jr_L + 2({}_jr_1)({}_jr_2)} = I_j \tag{12-95}$$

These equations constitute four equations in the four unknowns ${}_jr_1$, ${}_jr_2$, ${}_jE_1$, and ${}_jE_2$. The values of the remaining parameters in these equations are defined in Table 12-5 and are, therefore, known. By a rather routine but quite lengthy process of substitution and elimination, we obtain

$${}_jr_1 = \frac{{}_jr_L(1 - m_j)}{g_j({}_jr_L) + m_j} \tag{12-96}$$

$${}_jr_2 = \frac{{}_jr_L(1 + m_j)}{g_j({}_jr_L) - m_j} \tag{12-97}$$

$${}_jE_1 = \frac{(1 - m_j){}_jE_L - [1 + g_j({}_jr_L)]E_j - {}_jr_L(1 - m_j)I_j}{2[g_j({}_jr_L) + m_j]} \tag{12-98}$$

$${}_jE_2 = \frac{-(1 + m_j){}_jE_L + [1 + g_j({}_jr_L)]E_j - {}_jr_L(1 + m_j)I_j}{2[g_j({}_jr_L) - m_j]} \tag{12-99}$$

These relations completely specify ${}_jr_1$, ${}_jr_2$, ${}_jE_1$, and ${}_jE_2$ in terms of the given parameters.

It remains for us to determine the interval of definition of each segment j of R_A and R_B. To determine this, we solve for the relationships between the voltages across R_A and R_B in terms of v from Fig. 12-17. Thus we obtain

$$\begin{aligned} v_1 = v_3 = {} & \frac{{}_jr_1({}_jr_2 + {}_jr_L)}{\Delta} v \\ & + \frac{{}_jr_L[({}_jr_2)({}_jE_1) - ({}_jr_1)({}_jE_2)] - ({}_jr_1)({}_jr_2)({}_jE_L)}{\Delta} \end{aligned} \tag{12-100}$$

$$\begin{aligned} v_2 = v_4 = {} & \frac{{}_jr_2({}_jr_1 + {}_jr_L)}{\Delta} v \\ & + \frac{{}_jr_L[({}_jr_1)({}_jE_2) - ({}_jr_2)({}_jE_1)] + ({}_jr_1)({}_jr_2)({}_jE_L)}{\Delta} \end{aligned} \tag{12-101}$$

where Δ is as defined in Eq. (12-91). Solving for the driving-point v in terms of v_1 from Eq. (12-100), we obtain

$$v = \frac{\Delta}{{}_jr_1({}_jr_2 + {}_jr_L)} v_1 - \frac{{}_jr_L[({}_jr_2)({}_jE_1) - ({}_jr_1)({}_jE_2)] - ({}_jr_1)({}_jr_2)({}_jE_L)}{{}_jr_1({}_jr_2 + {}_jr_L)} \qquad (12\text{-}102)$$

Similarly solving Eq. (12-101) for v in terms of v_2, we obtain

$$v = \frac{\Delta}{{}_jr_2({}_jr_1 + {}_jr_L)} v_2 + \frac{{}_jr_L[({}_jr_2)({}_jE_1) - ({}_jr_1)({}_jE_2)] - ({}_jr_1)({}_jr_2)({}_jE_L)}{{}_jr_2({}_jr_1 + {}_jr_L)} \qquad (12\text{-}103)$$

Let us next impose on these equations the constraints given by Eqs. (12-86) and (12-87), namely,

$$a_j < v < b_j \qquad (12\text{-}104)$$

Substituting Eqs. (12-102) and (12-103) in place of v in Eq. (12-104), we obtain the following desired relationships:

Interval of definition for resistor R_A

$$\tfrac{1}{2}[(1 - m_j)a_j - E_j] < v_1 < \tfrac{1}{2}[(1 - m_j)b_j - E_j] \qquad (12\text{-}105)$$

Interval of definition for resistor R_B

$$\tfrac{1}{2}[(1 + m_j)a_j + E_j] < v_2 < \tfrac{1}{2}[(1 + m_j)b_j + E_j] \qquad (12\text{-}106)$$

Although the foregoing derivation is rather lengthy, the final synthesis equations are very simple and easy to apply. Indeed, only six equations are needed for the synthesis. The first four equations [Eqs. (12-96) to (12-99)] are used to calculate the slope and intercept of each segment of R_A and R_B, and the remaining two equations [Eqs. (12-105) and (12-106)] are used to calculate the interval of definition of the corresponding segment. An examination of Eqs. (12-96) and (12-97) shows that in order for the two resistors R_A and R_B to be monotonically increasing, we must have

$$-1 < m_j < 1 \qquad (12\text{-}107)$$

and

$$-g_j({}_jr_L) < m_j < g_j({}_jr_L) \qquad (12\text{-}108)$$

TABLE 12-6 *Pertinent information for specifying the resistor R_A.*

j	${}_j r_1$, Eq. (12-96)	${}_j E_1$, Eq. (12-98)	Interval of definition, Eq. (12-105)
1	0.632	-1.26	$(-\infty,-2)$
2	0.428	-1.5	$(-2,-1.5)$
3	0.333	-1.5	$(-1.5,-1)$
4	5.0	-8.5	$(-1,1.5)$
5	0.143	1.21	$(1.5,1.83)$
6	0.286	0.595	$(1.83,\infty)$

These two constraints correspond to conditions 2 and 3 of the lattice network synthesis theorem in Sec. 8-5. (Condition 1 of this theorem follows from the fact that R_A and R_B are monotonic.)

Let us now complete our synthesis of the specifications in Fig. 12-18. The pertinent parameters in Table 12-5 can now be substituted into Eqs. (12-96) to (12-99) to obtain the slope and intercepts of R_A and R_B. Similarly, substituting these parameters into Eqs. (12-105) and (12-106), we obtain the interval of definition for each segment. These results are summarized in Table 12-6 for the resistor R_A and in Table 12-7 for the resistor R_B. The reader should verify that these tables give precisely the same v_1-i_1 and v_2-i_2 curves obtained earlier in Fig. 8-17*g* and *h*.

Exercise 1: Derive the corresponding synthesis equations for the nonlinear voltage divider with R_1 as the nonlinear resistor.

Exercise 2: Derive the synthesis equations for realizing a prescribed DP plot and TC plot by a nonlinear voltage-divider network containing two nonlinear resistors R_1 and R_2.

Exercise 3: Derive Eqs. (12-89) through (12-106) and verify results given in Tables 12-6 and 12-7.

Exercise 4: Describe the step-by-step procedure for deriving the synthesis equations for realizing a jointly prescribed DP plot, load characteristics, and the following type of TC plots: (*a*) i_o versus i; (*b*) v_o versus i; (*c*) i_o versus v.

TABLE 12-7 *Pertinent information for specifying the resistor R_B.*

j	${}_j r_2$, Eq. (12-97)	${}_j E_2$, Eq. (12-99)	Interval of definition, Eq. (12-106)
1	2.77	0.23	$(-\infty,-3)$
2	9.0	7.5	$(-3,-1.5)$
3	3.0	1.5	$(-1.5,0)$
4	0.20	0.10	$(0,0.5)$
5	7.0	-13.5	$(0.5,2.83)$
6	14.0	-29.83	$(2.83,\infty)$

Exercise 5: In many practical cases the load and the DP plot are linear. Simplify the synthesis equations for this case, and show how the resistance corresponding to the load and the DP plot may be chosen in connection with the given TC plot so that both R_A and R_B will be monotonic.

12-7 SUMMARY

Basic philosophy Each network must be recast into the *canonic form* so that all nonlinear elements are two-terminal resistors characterized by piecewise-linear v-i curves. Any resistive network may be recast into this form by substituting each n-terminal ($n > 2$) resistor by a network model in the canonic form.

Iterative linear network Each nonlinear resistor characterized by a piecewise-linear v-i curve is equivalent to an iterative Thévenin equivalent circuit or to an iterative Norton equivalent circuit. A nonlinear network in the canonic form is said to be an iterative linear network if all nonlinear resistors in the network have been replaced by their iterative Thévenin or Norton equivalent circuits.

The iterative piecewise-linear method Given a nonlinear network in the canonic form, we first obtain its associated iterative linear network. This network is linear and can be analyzed by a number of available techniques. Any voltage or current expressions derived from this network will be in terms of the iterative parameters associated with the nonlinear resistors. This part of the analysis is no different from the analysis of any other linear network.

The second part of the analysis consists of a systematic iteration over all possible segment combinations. No new concept is involved here other than an efficient bookkeeping system. Such a task is ideally suited for a digital computer. The iterative piecewise-linear method is the most general method available. It can be used to find the operating points, the DP plots, and the TC plots of any resistive nonlinear network in the canonic form.

Computer network analysis Two versions of general computer programs are available:

1. Semiautomatic general analysis program. The user will perform the first part of the iterative piecewise-linear method by deriving all pertinent equations needed by the program. The iteration part of this method is carried out by the computer. The disadvantage of this approach is that for complicated networks, even the linear analysis part could involve a lot of work. The advantage of this approach is that once the expressions are derived for a particular network, such as a Schmitt trigger, the program can be considered as completely automatic for *this*

network. This not only reduces computer time, but also allows the effect of parameter variations to be studied. Moreover, because the program is not completely automatic, it can be kept to a small size.

2. Completely automatic Meca program. In this version the user does not have to derive any equation. The network topology and element interconnections are given directly as the input to the computer. The computer essentially derives the pertinent equations and carries out the iterations. This program is very powerful because it applies to multivalued networks as well. It is completely user-oriented. The only disadvantage is that the size of the program is very large and hence requires a large computer for its implementation. This is the price that one pays for achieving complete automation.

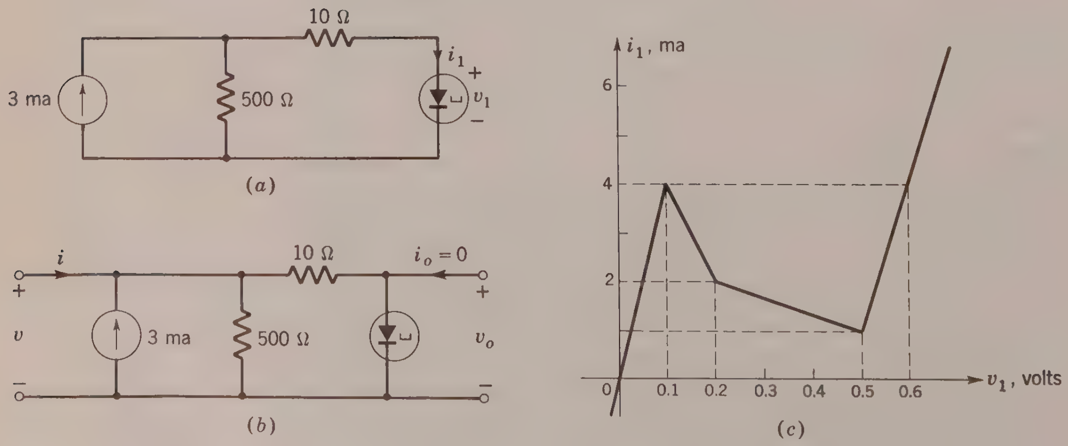
Synthesis by the iterative piecewise-linear method Most of the synthesis methods described in Chap. 8 can be derived by the iterative piecewise-linear method, which offers two advantages over the graphical methods:

1. It makes possible deriving synthesis equations so that the restrictions of the circuit can be easily determined.
2. It is well-suited for digital computer programming.

PROBLEMS

12-1 Consider the networks shown in Fig. P12-1a and b, where the tunnel diode v_1-i_1 curve is given in Fig. P12-1c.

Fig. P12-1.



(a) Using the iterative Thévenin equivalent circuit, find the operating point of the network in Fig. P12-1*a* and the DP plot and TC plot of the network in Fig. P12-1*b*.
(b) Repeat (*a*) using the iterative Norton equivalent circuit.
(c) Check the above solutions by the graphical method.

12-2 Consider the network shown in Fig. P12-2*a*, where the v_1-i_1 curve is shown in Fig. P12-2*b*.
(a) Find the operating point when a current source $i = 3$ ma is connected across the driving-point terminals.
(b) Find the v_1-vs.-v TC plot.
(c) Find the i_1-vs.-i TC plot.

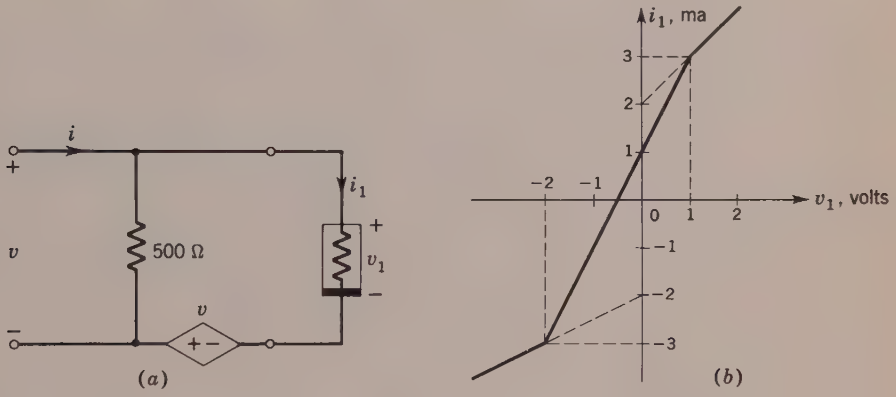
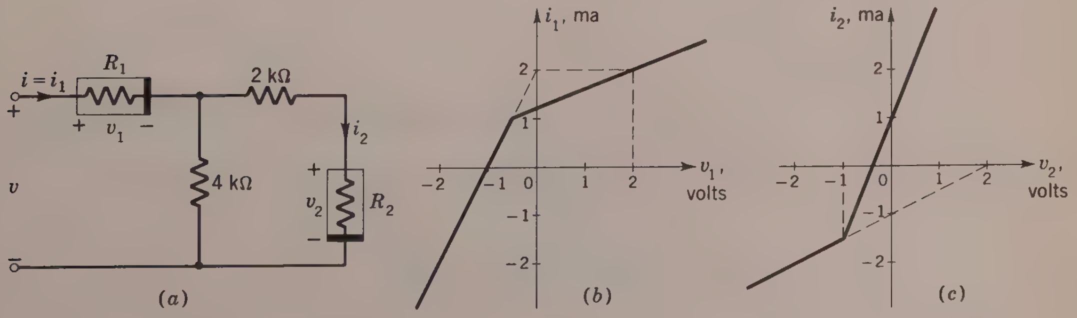


Fig. P12-2.

12-3 Consider the network shown in Fig. P12-3*a*, where the v_1-i_1 and the v_2-i_2 curves of R_1 and R_2 are shown in Fig. P12-3*b* and *c*.
(a) Find the operating point when a voltage source $v = 3$ volts is connected across the driving-point terminals.
(b) Find the *v*-*vs.*-*i* DP plot.
(c) Find the v_2-vs.-v TC plot.
(d) Find the v_1-vs.-v TC plot.

Fig. P12-3.



12-4 Consider the network shown in Fig. P12-4 where the v-i curves of the nonlinear resistors R_1 and R_2 are given by Fig. 12-5*b* and *c* in the text.

(a) Find the operating point when a voltage source $v = 5$ volts is connected across the driving-point terminals.

(b) Find the v-vs.-i DP plot.

(c) Find the v_2-vs.-v TC plot.

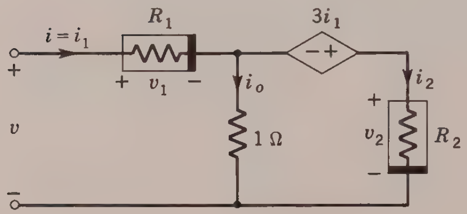


Fig. P12-4.

12-5 Consider the network shown in Fig. P12-5*a*, where the v_1-i_1 and the v_2-i_2 curves of the nonlinear resistors are shown in Fig. P12-5*b* and *c*, respectively.

(a) Find the operating point when a 0.5-ma current source is connected across the driving-point terminals.

(b) Find the v-vs.-i DP plot.

(c) Find the v_2-vs.-v TC plot.

(d) Find the v_2-vs.-i TC plot.

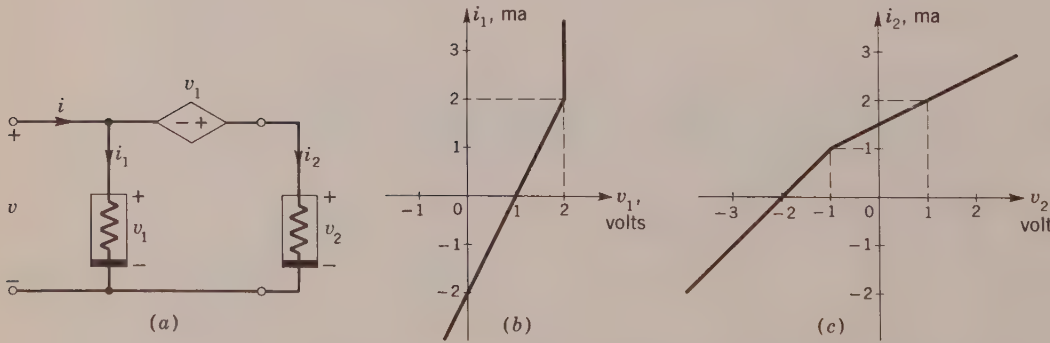


Fig. P12-5.

12-6 Consider the network shown in Fig. P12-6*a*, where the v_1-i_1 and the v_2-i_2 curves of the nonlinear resistors are shown in Fig. P12-6*b* and *c*, respectively.

(a) Find the operating point when a 5-volt voltage source is connected across the driving-point terminals.

(b) Find the v-vs.-i DP plot.

(c) Find the v_2-vs.-v TC plot.

(d) Find the i_2-vs.-i TC plot.

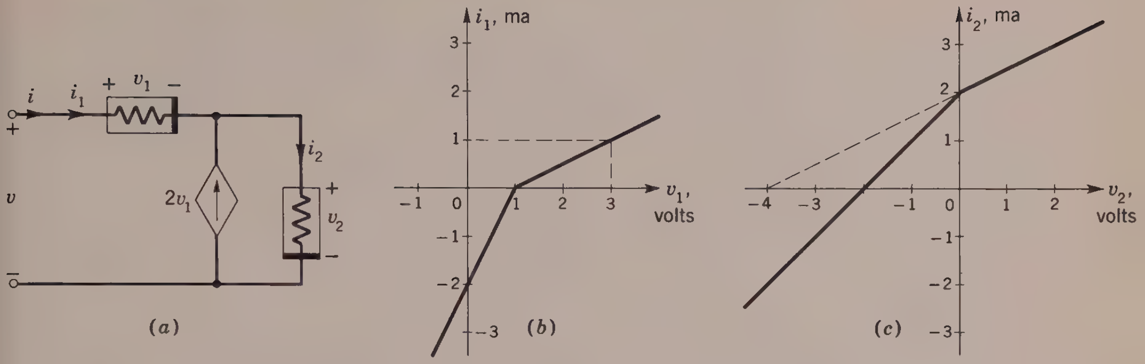


Fig. P12-6.

12-7 When a network contains only ideal diodes as the nonlinear elements, it is sometimes easier to analyze a series of reduced linear networks obtained by replacing the ideal diodes by either open circuits or short circuits corresponding to each segment combination. If there are n ideal diodes, a total of 2^n reduced linear networks must be analyzed. This is called the *method of assumed states.*

(a) Find the DP plot of the network shown in Fig. P12-7 by the iterative piecewise-linear method.

(b) Repeat (*a*) by the method of assumed states.

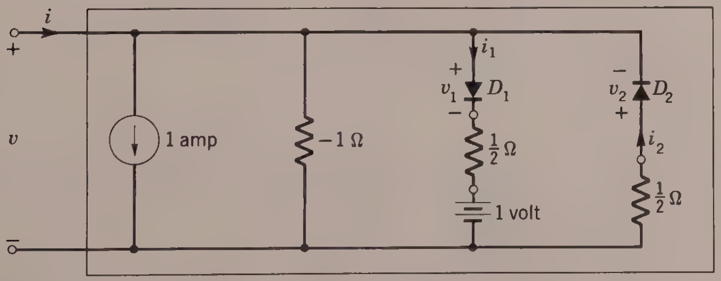


Fig. P12-7.

12-8 Do Prob. 12-7 for the ideal diode network shown in Fig. P12-8.

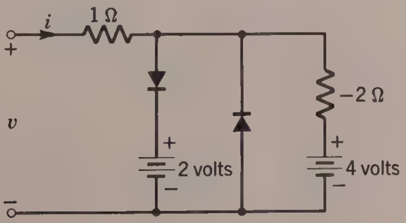


Fig. P12-8.

12-9 Do Prob. 12-7 for the ideal diode network shown in Fig. P12-9.

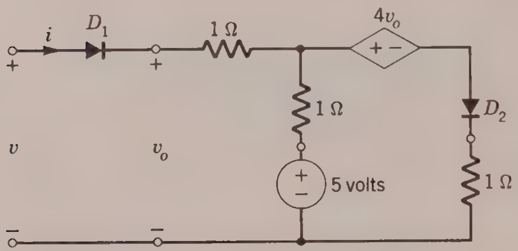


Fig. P12-9.

12-10 Consider the transistor circuit shown in Fig. P12-10.

(a) Find the operating point corresponding to a driving-point voltage $v = -0.4$ volt.

(b) Find the v-vs.-i DP plot.

(c) Find the v_o-vs.-v TC plot.

(d) Find the power-transfer plot between the load resistor R_L and the voltage source.

NOTE: Use the *p-n-p* transistor model B_E.

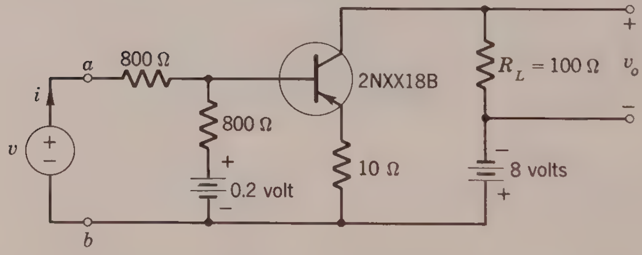


Fig. P12-10.

12-11 Consider the transistor circuit shown in Fig. P12-11.

(a) Find the operating point corresponding to a driving-point voltage $v = 0.4$ volt.

(b) Find the v-vs.-i DP plot.

(c) Find the v_o-vs.-v TC plot.

(d) Find the power-transfer plot between the load resistor R_L and the voltage source.

NOTE: Use the *n-p-n* transistor model B_B.

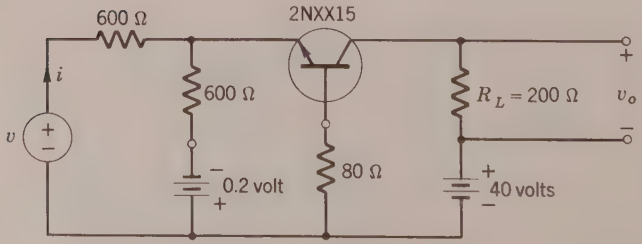


Fig. P12-11.

12-12 Consider the triode network shown in Fig. P12-12. Using the triode model C with $r_p = 3.3\ \text{k}\Omega$, $\mu = 20$, and $r_g = 250\ \Omega$, find the v-vs.-i DP plot and the v_o-vs.-v TC plot.

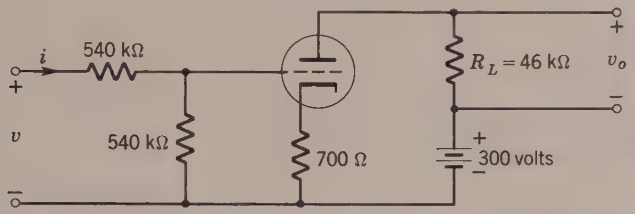


Fig. P12-12.

12-13 The network shown in Fig. P12-13 is the vacuum-tube version of the Schmitt trigger circuit. Using the triode model C with $r_p = 5\ \text{k}\Omega$, $\mu = 70$, and $r_g = 250\ \Omega$, find the v_o-vs.-v TC plot.

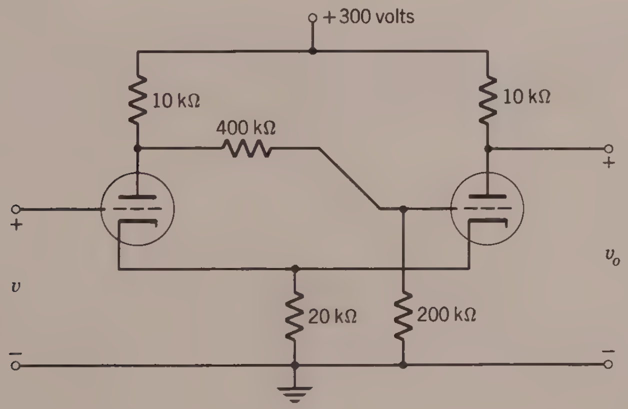


Fig. P12-13.

12-14 A transistorized Eccles-Jordan circuit is shown in Fig. P12-14. Using the *n-p-n* transistor model B_E, find the operating points of this network.

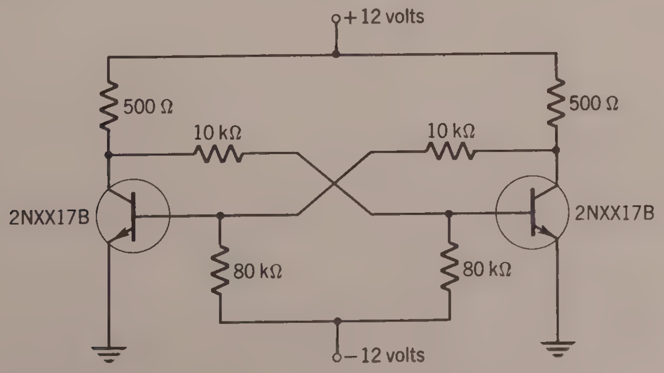


Fig. P12-14.

12-15 Use the piecewise-linear method to synthesize the TC plots shown in Fig. P12-15*a* and *b* by a nonlinear voltage divider.

(a) Choose $R_1 = 100\ \Omega$.
(b) Choose $R_2 = 100\ \Omega$.

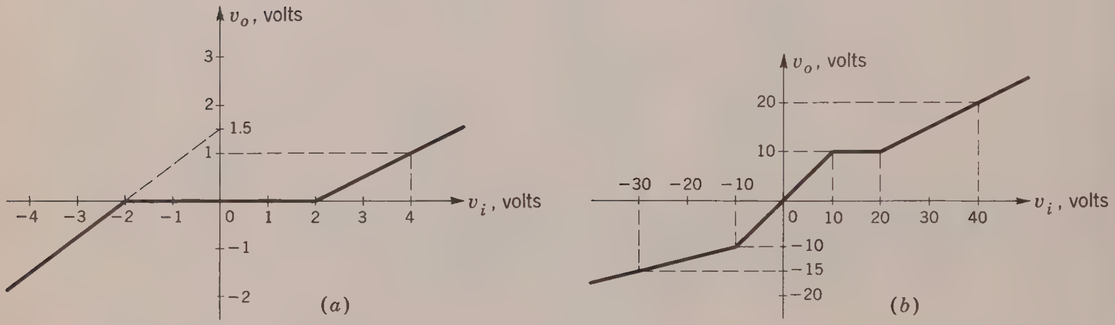


Fig. P12-15.

12-16 Use the piecewise-linear method to synthesize the jointly prescribed TC plot, load v_L-i_L curve, and DP plot shown in Fig. P8-18*a* to *c* in Chap. 8.

part 3

DYNAMIC NONLINEAR NETWORKS

13 BASIC CONCEPTS OF DYNAMIC NONLINEAR NETWORKS

13-1 CLASSIFICATION OF DYNAMIC NONLINEAR NETWORKS

By the definition in Sec. 4-1-2, dynamic nonlinear networks represent the class of all nonlinear networks other than resistive networks. In view of the negative nature of this definition, the classes of resistive and dynamic networks together constitute the universe of all networks, because a network is either resistive or dynamic. If we represent the universe of all networks symbolically by the set of points U (where each point represents a network) as shown in Fig. 13-1*a*, then the heavily shaded portion of U represents the class of all resistive networks, and the lightly shaded portion of U represents the class of all dynamic networks. Since the class of dynamic networks is so much larger than the class of resistive networks, it would be too ambitious for us to formulate a general theory that is applicable to the solution of all dynamic networks. After all, it took us eight chapters just to formulate a general theory to handle resistive networks. The logical approach, therefore, is to break up the huge class of dynamic networks into various subclasses as shown in Fig. 13-1*b* and attempt to formulate a general theory to handle each subclass.

13-1-1 BASIS OF CLASSIFICATION

What principle should be used to classify dynamic networks? Obviously, dynamic networks having some common properties or features must be grouped in the same class. Since the essential feature of any dynamic network is the presence of at least one energy-storage element (namely, an inductor or a capacitor), it seems reasonable to classify each subclass of dynamic networks in terms of the number of energy-storage elements present in the networks.

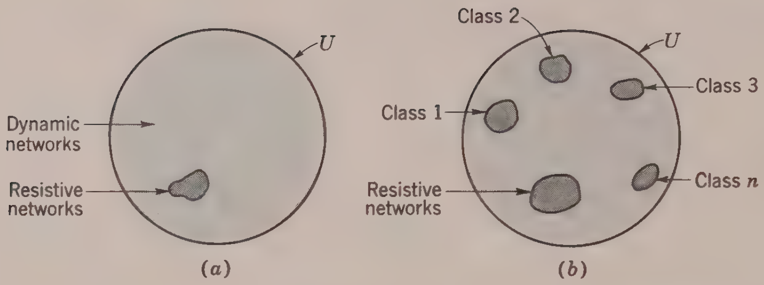


Fig. 13-1. Symbolic representation of the universe of all networks and their classifications.

However, this is not very satisfactory because certain dynamic networks containing different numbers of energy-storage elements can be analyzed using exactly the same principles, and hence there is no reason to separate them. For example, consider the simple one-energy-storage-element network N_a shown in Fig. 13-2*a* and the two-energy-storage-element network N_b shown in Fig. 13-2*b*. The equation of motion for N_a is readily seen to be

$$RC\frac{dv_1}{dt} + v_1 = \sin t \tag{13-1}$$

Similarly, the equation of motion for N_b is

$$R(C_1 + C_2)\frac{dv_1}{dt} + v_1 = \sin t \tag{13-2}$$

Observe that Eqs. (13-1) and (13-2) are similar in form, and their respective solutions can be found using identical procedures. It is more natural, therefore, to consider N_a and N_b as belonging to the same subclass rather than to separate them, as would be the case if we adopt as the basis of classification the number of energy-storage elements.

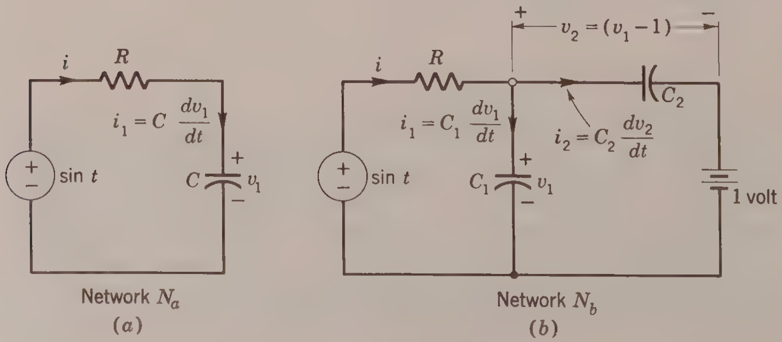


Fig. 13-2. Two networks containing different numbers of energy-storage elements but similar equations of motion.

Since one of the most basic problems in dynamic networks is to find the solution of the equations of motion consisting of a system of nonlinear differential equations, it is more appropriate to classify dynamic networks according to the "complexity" of their system of differential equations of motion. It is well known that the solution to any system of differential equations can be found only to within a number of arbitrary constants $k_1, k_2, \ldots, k_n$. In order to determine the n arbitrary constants, we must specify n independent "initial conditions." In other words, there are as many arbitrary constants as there are independent initial conditions.[1]

Two systems of differential equations requiring different numbers of initial conditions are usually solved by quite different methods.[2] Since our ultimate objective is to solve these equations, a more meaningful basis for the classification of dynamic networks can be stated in terms of the number of *independent* initial conditions that must be specified in order to uniquely solve for the solution of the network. In view of the above considerations, we are now ready to state a definition which serves as a basis for classifying dynamic networks.

13-1-2 DEFINITION OF ORDER OF COMPLEXITY

We define the *order of complexity* of a dynamic network to be the maximum number n of independent initial conditions that can be specified in terms of the electrical variables in the network.[3] For convenience we shall refer to N as a *first-order network* if $n = 1$ and as a *second-order network* if $n = 2$. Since $n \geq 1$ for any dynamic network, we might, for the sake of completeness, refer to any resistive network as a *zero-order network*. It is important to observe that the above definition requires that the number of initial conditions be independent of one another. The word "independent" means that none of the specified initial conditions can be derived from the rest. As an illustration of the above definition, consider again the network N_a shown in Fig. 13-2*a*. Since N_a contains only one energy-storage element, we can specify only one initial condition, namely, the voltage $v_1(t_0)$ across the capacitor C at some arbitrary time t_0; hence, N_a is a first-order network. Consider next the network N_b shown in Fig. 13-2*b*. Since N_b contains two energy-storage elements, it appears at first sight that we can specify two initial conditions, namely, the voltage $v_1(t_0)$ across capacitor C_1 and the voltage $v_2(t_0)$ across capacitor C_2 at some time t_0. However, since $v_2 = v_1 - 1$, the two initial conditions are *not* independent because once $v_1(t_0)$ is specified, $v_2(t_0)$ is constrained to be $v_1(t_0) - 1$, and hence $v_2(t_0)$ cannot be specified arbitrarily. Therefore N_b is a first-order network because we can specify only one *independent* initial condition. From the theory of differential equations, it is known that a system of n differential equations in the normal form[1] requires exactly n initial conditions for its solution. Therefore the order of a dynamic network is known once the equations of motion of the network have been reduced to the normal form. In particular, the number of state variables is equal to the order of complexity of the network.

---

[1] A set of initial conditions is said to be independent if its values can be arbitrarily chosen.

[2] The only general method which does not depend on the number of differential equations is the *numerical integration technique* discussed in Secs. 4-7 and 20-2. However, this method is not particularly useful for studying the qualitative properties of nonlinear networks. Because of this limitation, only qualitative methods are emphasized in the following chapters.

[3] As will be obvious shortly, the order of complexity of a dynamic network is equal to the *minimum* number of initial conditions that *must* be specified in order to specify completely the behavior of the network.

13-2 ORDER OF COMPLEXITY OF DYNAMIC NETWORKS

As will be obvious in the sequel, our classification of dynamic networks according to their order of complexity is very useful because each subclass can usually be handled by some unified principles and techniques. Since the theory for different orders of networks is in general quite different, it is necessary for us to determine first the order of complexity of a given network before we know which theory to apply. One method for obtaining this information is, of course, to write the equations of motion of the network in the normal form. We shall develop in this section a simpler technique for determining the order of complexity of any dynamic *RLC* network by inspection, i.e., without writing down any equation.[2] In order to do this, it is important for us to obtain a deeper understanding of why initial conditions are necessary from the network's point of view, and to understand which electrical variables qualify as an appropriate set of initial conditions.

13-2-1 SIGNIFICANCE OF THE INITIAL CONDITION

From the mathematical point of view, initial conditions are introduced as a "gimmick" for determining the values of the arbitrary constants associated with the solution to a system of differential equations. From the network's point of view, initial conditions are introduced because of our ignorance or incomplete knowledge of the past history of the excitations that have been applied to the network. In order to understand the above reason, let us consider an arbitrary capacitor C_j of an arbitrary network N. Suppose we want to find the charge $q_j(t)$ of this capacitor at time t, namely,

$$q_j(t) = \int_{-\infty}^{t} i_j(\tau)\, d\tau \tag{13-3}$$

---

[1] Review Sec. 4-6 for the significance of normal form and state variables.

[2] By *RLC* network we mean a network containing only two-terminal resistors, inductors, capacitors, and independent sources. No controlled sources, ideal transformers, gyrators, etc., are allowed.

From Eq. (13-3) it is clear that $q_j(t)$ can be found only if we know the exact waveform of the capacitor current $i_j(t)$ from $t = -\infty$ up to the present time t, that is, from the time the capacitor was manufactured. However, any network N is constructed at some finite time in the past, and in any physical network the excitations are applied at some time, say, $t = t_0$. Hence, we would usually have information on the excitation waveforms only for $t \geq t_0$, since by assumption, this is the time when the excitations are applied. To be specific, suppose we know the capacitor current $i_j(t)$ for $t \geq t_0$, where $t = t_0$ is the time we closed the switch which connects the current source $i_j(t)$ across the capacitor C_j. Unfortunately, this information is not sufficient for us to determine $q_j(t)$ from Eq. (13-3), because we need to know $i_j(t)$ during $-\infty < t < t_0$ as well. Although our excitation was applied at $t = t_0$, we have no right to assume that $i_j(t) = 0$ for $t < t_0$ because, sometime in the remote past, $i_j(t)$ may not have been zero. For example, the capacitor may have been charged prior to $t = t_0$ either by someone else or by a number of random sources such as lightning or atmospheric noise.

Our ignorance of the past history of $i_j(t)$ prevents us from determining $q_j(t)$. However, since the excitations are applied at $t = t_0$, we are really interested to know only $q_j(t)$ for $t \geq t_0$. Let us rewrite Eq. (13-3) in the form

$$q_j(t) = \int_{-\infty}^{t_0} i_j(\tau)\, d\tau + \int_{t_0}^{t} i_j(\tau)\, d\tau \tag{13-4}$$

The second integral can be found because we know $i_j(t)$ for $t \geq t_0$. It is the first integral which is causing the trouble. Observe, however, that at $t = t_0$, Eq. (13-3) becomes

$$q_j(t_0) = \int_{-\infty}^{t_0} i_j(\tau)\, d\tau \tag{13-5}$$

where $q_j(t_0)$ is the total charge in the capacitor at $t = t_0$. Substituting Eq. (13-5) into Eq. (13-4), we obtain

$$q_j(t) = q_j(t_0) + \int_{t_0}^{t} i_j(\tau)\, d\tau \tag{13-6}$$

where $t \geq t_0$.

Equation (13-6) tells us that, provided we are interested only in knowing $q_j(t)$ for $t \geq t_0$, it is not necessary to know the entire past history of $i_j(t)$ for $t < t_0$. Instead, we need to know only the value of the charge q_j in the capacitor at the initial time t_0. The value $q_j(t_0)$ is called the *initial condition.* In other words, we have shown that in order to determine the charge $q_j(t)$ in a capacitor, it is necessary to know not only the excitation but also the initial condition. It should now be clear that the single piece of information $q_j(t_0)$ is used in lieu of the entire past history of $i_j(t)$.

Let us next recall that a capacitor is characterized by a curve in the v-q plane, and if we know $v(t)$, we can find $q(t)$, and vice versa.[1] Since it is necessary to know the initial condition $q(t_0)$ in order to find $q(t)$ for $t \geq t_0$, it follows that it is necessary to know $v(t_0)$ in order to find $v(t)$ for $t \geq t_0$. However, since given $v(t_0)$, we can find $q(t_0)$ from any voltage-controlled v-q curve, or given $q(t_0)$, we can find $v(t_0)$ from any charge-controlled v-q curve, it is clear that in order to find $q(t)$ and $v(t)$ across a capacitor, it is sufficient to specify an initial condition either in terms of the capacitor charge $q(t_0)$ or in terms of the capacitor voltage $v(t_0)$ at the initial time t_0.

Finally, an examination of Eq. (13-6) shows that specifying the capacitor current $i_j(t_0)$ at t_0 would not do any good because one cannot determine $q_j(t_0)$ from this information alone. We conclude, therefore, that the current in a capacitor is not an appropriate initial condition.

By exact dual arguments, we find that for an inductor

$$\varphi_j(t) = \varphi_j(t_0) + \int_{t_0}^{t} v_j(\tau)\, d\tau \tag{13-7}$$

where $t \geq t_0$.

In order to find the flux linkages $\varphi_j(t)$ for $t \geq t_0$, it is necessary to specify the initial condition $\varphi_j(t_0)$ at t_0. Moreover, since an inductor is characterized by a curve in the i-φ plane, it follows that we can specify either the flux linkage $\varphi(t_0)$ or the inductor current $i(t_0)$ at t_0 as the initial condition. An examination of Eq. (13-7) shows that the voltage across an inductor is not an appropriate initial condition.

The above results are important enough to be stated in the form of a theorem.

THEOREM ON THE CHOICE OF APPROPRIATE INITIAL CONDITIONS

An appropriate set of initial conditions that can be specified for a dynamic network consists of the value of the charge q or voltage v across a capacitor and the value of the flux linkage φ or current i in an inductor at some arbitrary initial time $t = t_0$ (which is usually chosen to be zero). The capacitor current and the inductor voltage are not appropriate initial conditions.

---

[1] To be exact, given $v(t)$, we can find $q(t)$ uniquely if the v-q curve is voltage-controlled. Similarly, given $q(t)$, we can find $v(t)$ uniquely if the v-q curve is charge-controlled.

This theorem specifies a fundamental set of initial conditions that are commonly used. There are other possibilities which are just as valid. Indeed, any linearly independent combination of capacitor voltages and inductor currents would qualify. For example, any set of initial conditions from which the capacitor voltages and inductor currents can be calculated is valid.

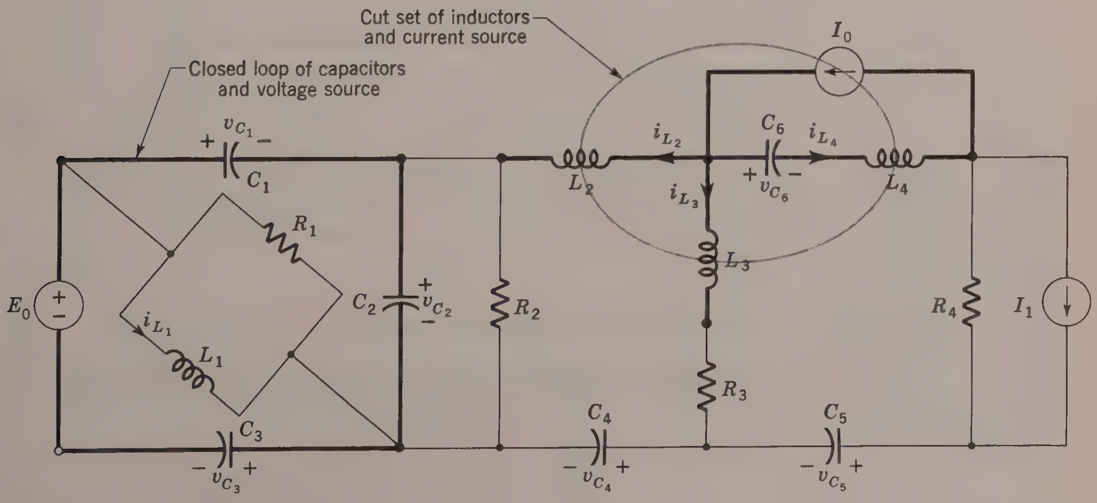
13-2-2 DETERMINATION OF ORDER OF COMPLEXITY BY INSPECTION

We have already seen that the order of complexity of a dynamic network may not be equal to the number of energy-storage elements because some initial conditions may not be independently specified. In order to diagnose the source of "dependency," let us consider the more complicated network N in Fig. 13-3.

Since N contains 10 energy-storage elements (six capacitors and four inductors), it appears that we can specify 10 initial conditions, v_{C_1}, v_{C_2}, v_{C_3}, v_{C_4}, v_{C_5}, v_{C_6}, i_{L_1}, i_{L_2}, i_{L_3}, and i_{L_4}. However, a more careful inspection of the network shows that not all these initial conditions are independent. For example, the loop consisting of capacitors C_1, C_2, and C_3 and voltage source E_0 imposes a constraint due to KVL; thus

$$v_{C_1} + v_{C_2} + v_{C_3} = E_0 \tag{13-8}$$

Fig. 13-3. An example of the two possible sources of dependent initial conditions, namely, a loop of capacitors and voltage sources, and a cut set of inductors and current sources.



This equation implies that only two of the three initial conditions v_{C_1}, v_{C_2}, and v_{C_3} can be specified arbitrarily. We conclude that although there are six capacitors, only five capacitor voltages are independent. Similarly, the cut set consisting of inductors L_2, L_3, and L_4 and current source I_0 imposes a constraint due to KCL; thus

$$i_{L_2} + i_{L_3} + i_{L_4} = I_0 \tag{13-9}$$

Equation (13-9) implies that only two of the three initial conditions i_{L_2}, i_{L_3}, and i_{L_4} can be specified arbitrarily. Hence we conclude that although there are four inductors, only three inductor currents are independent. The maximum number of independent initial conditions that can be specified is therefore equal to $5 + 3 = 8$, and hence network N in Fig. 13-3 is an eighth-order network.

Based on the above example, it is clear that a dependency exists whenever it is possible to write a constraint involving only capacitor voltages and voltage sources; therefore, we must subtract one initial condition from the total number of energy-storage elements. Similarly, it is clear that a dependency exists whenever it is possible to write a constraint involving only inductor currents and current sources; therefore, we must likewise subtract one initial condition from the total number of energy-storage elements. The first constraint involving only capacitor voltages and voltage sources occurs if and only if there exists a loop in the network containing *only* capacitors and independent voltage sources. Similarly, the second constraint involving only inductor currents and current sources occurs if and only if there exists a cut set in the network containing *only* inductors and independent current sources. If we count the total number of independent capacitor-voltage-source-only loops and the total number of independent inductor-current-source-only cut sets and subtract these from the total number of energy-storage elements, we obtain the maximum number of independent initial conditions.[1] The above observations are important enough to be stated in the form of a theorem.

THEOREM ON ORDER OF COMPLEXITY OF DYNAMIC *RLC* NETWORKS

The order of complexity of any dynamic network N is given by

$$n = b_{LC} - n_C - n_L \tag{13-10}$$

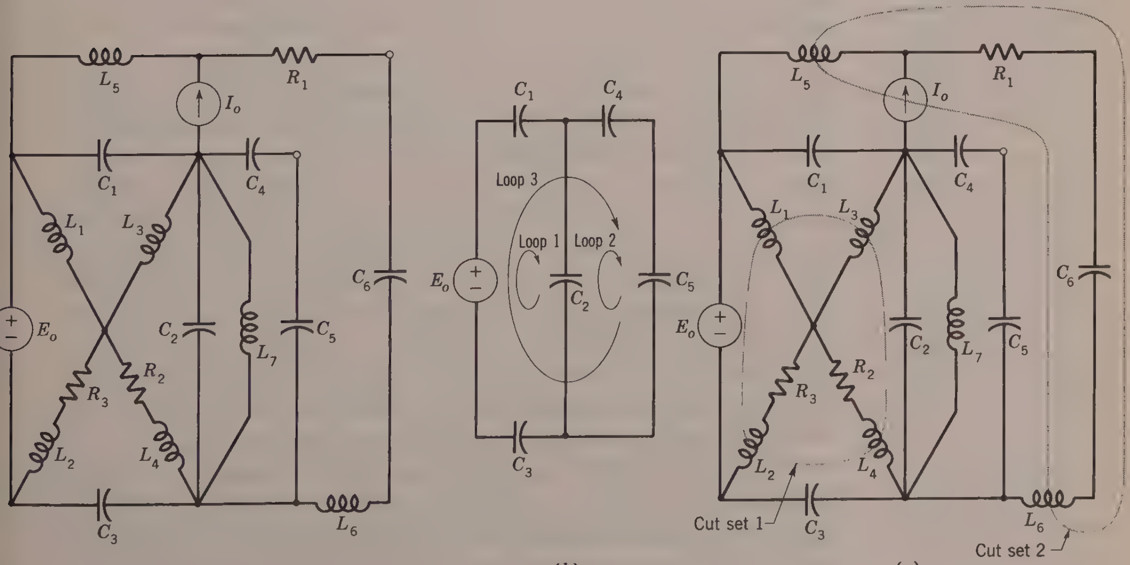
---

[1] By independent we mean that no constraint is a linear combination of the remaining constraints. Clearly, a set of loops is independent if each loop contains at least one capacitor not contained in the other loops. Similarly, a set of cut sets is independent if each cut set contains at least one inductor not contained in the other cut sets.

where b_{LC} = total number of energy-storage elements in N
n_C = total number of *independent* loops containing only capacitors and possibly voltage sources
n_L = total number of *independent* cut sets containing only inductors and possibly current sources

Applying the above theorem to the network shown in Fig. 13-3, we find that $b_{LC}=10$, $n_C=1$, and $n_L=1$; hence, $n=10-1-1=8$, as expected. For a more complicated example, consider the network shown in Fig. 13-4*a*. It appears at first sight that there are three loops containing only capacitors and voltage sources as shown in Fig. 13-4*b*. However, only two of the three loops are independent because the KVL equation written for the third loop can be derived from the KVL equations written for loops 1 and 2. Therefore, we conclude that $n_C = 2$. From Fig. 13-4*b* it is clear that in order for a loop to be independent with respect to a set of previously chosen independent loops, it must contain at least one new capacitor not contained in the previous loops. In order to determine n_L, we find that there are two independent cut sets containing only inductors and current sources as shown in Fig. 13-4*c*; hence $n_L = 2$. Applying the above theorem, we find that $n = 13 - 2 - 2 = 9$ and the network in Fig. 13-4 is a ninth-order network.

Fig. 13-4. A network containing dependent loops of capacitor and voltage sources.



(*a*) (*b*) (*c*)

It should now be clear that this theorem allows us to determine by inspection the order of complexity of any dynamic *RLC* network. We need determine only n_C and n_L. To determine n_C, we may replace all resistors, inductors, and current sources in N by open circuits, thus leaving a much simpler subnetwork containing only capacitors and voltage sources, and n_C can be found by inspection. Similarly, to determine n_L, we may replace all resistors, capacitors, and voltage sources in N by short circuits, thus leaving a much simpler subnetwork containing only inductors and current sources, and n_L can be found by inspection.[1]

13-2-3 SOME SUBTLE POINTS CONCERNING INITIAL CONDITIONS

There are three questions concerning initial conditions which are of considerable theoretical interest. Since a complete answer to these questions is beyond the scope of this book, we shall only briefly mention them here. The first question is whether the theorem on the order of complexity of dynamic *RLC* networks can be generalized to include controlled sources in the network. The answer is yes, provided additional restrictions are imposed on the network topology. Fortunately, for most practical networks of interest these additional restrictions are usually satisfied automatically, and hence this theorem is still applicable. In general, however, we must watch for the possibility of some additional constraints that may be introduced by the controlled sources, thereby reducing further the order of complexity of the network.

A second question of interest concerns the definition of equivalent dynamic networks. Clearly, the initial conditions must be included in any such definitions. Since a meaningful discussion of this topic entails some knowledge of differential equations, which we do not assume in this book, we shall just point out that the concept of equivalent dynamic networks is not as practical as for resistive networks because the need for including initial conditions complicates the problem considerably. This complication tends to obliterate the advantage of introducing equivalent networks, that is, simplifying the analysis problem.

The third question concerns the effect on the order of complexity of the original network caused by the introduction of an equivalent network. For example, if we have two capacitors in series, can we replace them by an equivalent capacitor? The answer is yes, if we take the initial condition of each capacitor into consideration. In so doing, we have reduced the order of complexity by 1; but we have also lost some information in the process. Since we shall not encounter these subtle points in the remainder of this book, or in most practical networks, let us proceed to the next section.

---

[1] For more complicated networks, it may be desirable to apply the topological techniques presented in Sec. 4-3 for finding n_C and n_L. For example, applying the independent KVL equation criteria in Sec. 4-3-1, we obtain the relationship $n_C = b' - n' + 1$, where b' and n' represent the total number of branches and nodes, respectively, in the *capacitor-voltage-source subnetwork.* It is important to note that only capacitors and voltage sources which form closed loops with at least one other capacitor or voltage source are considered to be part of the subnetwork. For example, the capacitor C_6 in Fig. 13-4*a* is not counted as part of the subnetwork. Observe that in this case, $b' = 6$ and $n' = 5$, and hence $n_C = 2$, as expected. By a dual procedure, we obtain from the independent KCL equation criteria in Sec. 4-3-2 the relationship $n_L = n' - 1$, where n' represents the total number of nodes in the *inductor-current-source subnetwork.*

Exercise 1: Define state variables, order of complexity, and number of independent initial conditions. What are the relationships between these three concepts?

Exercise 2: State a simple criterion for guaranteeing that the order of complexity is equal to the number of energy-storage elements.

Exercise 3: Energy-storage elements are said to possess memories. Explain why.

Exercise 4: Find a simple example showing how a controlled source may introduce an additional constraint not considered in the theorem on the order of complexity of dynamic RLC networks.

Exercise 5: (*a*) Show that under the assumption of zero initial conditions, a two-terminal black box containing only inductors (capacitors) can be replaced by an equivalent inductor (capacitor). (*b*) Show that the series- and parallel-combination techniques for resistive networks are applicable here. State the corresponding procedure in terms of the i-φ (v-q) curves.

Exercise 6: Generalize the theorem on the order of complexity of dynamic RLC networks to include linear n-terminal resistors such as ideal transformers, gyrators, rotators, and reflectors but not controlled sources.

13-3 BASIC PRINCIPLES FOR ANALYZING DYNAMIC NETWORKS

One of the most basic principles for solving a new problem in any discipline is to resolve the problem into a number of simpler interrelated subproblems whose solutions can be obtained by known techniques. Using this principle, we shall resolve a dynamic network N into an interconnection of two subnetworks, namely, an *energy-storage subnetwork* N_1 and a *resistive subnetwork* N_2. The energy-storage subnetwork N_1 is obtained by extracting all energy-storage elements, two-terminal, three-terminal, . . . , n-terminal inductors and capacitors, from the dynamic network N. What remains is the resistive subnetwork N_2 containing all linear and nonlinear resistors, independent and controlled sources, ideal transformers, gyrators, rotators, NIC, etc.

For the purpose of this book we shall consider only two-terminal inductors and two-terminal capacitors as the energy-storage elements. Under this condition the energy-storage subnetwork N_1 can be redrawn explicitly as shown in Fig. 13-5*b*. In this case, the resistive subnetwork N_2 becomes an n-port network.

To analyze the dynamic network shown in Fig. 13-5*b*, the first step is to obtain the equations of motion in the normal form in terms of the state variables i_{L_j} or φ_{L_j} of the inductors and v_{C_j} or q_{C_j} of the capacitors. The next step is to solve these differential equations. This is the most difficult task in the analysis and

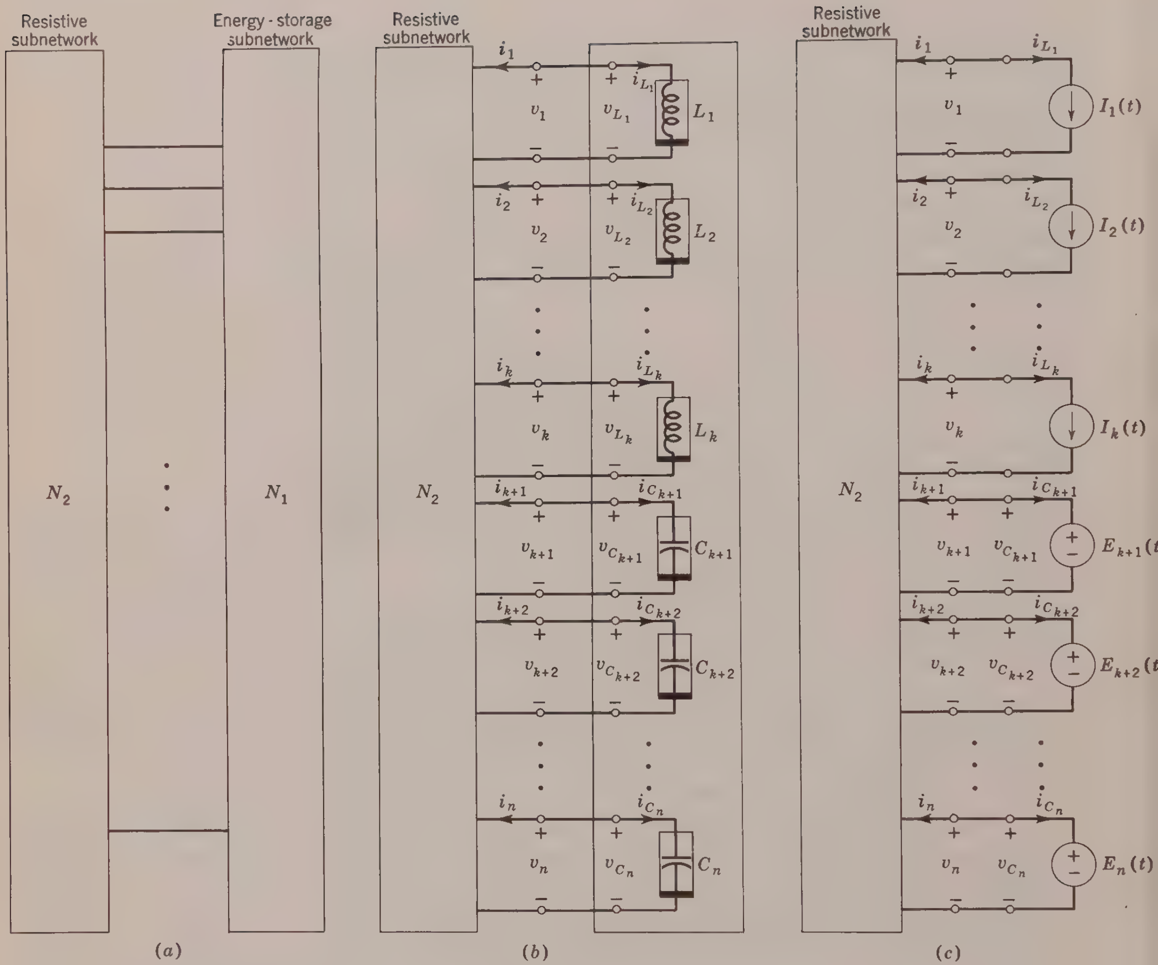


Fig. 13-5. Any dynamic network can be resolved into an energy-storage subnetwork and a resistive subnetwork. The energy-storage subnetwork can be replaced by ac sources for the purpose of analyzing the resistive subnetwork.

will be considered in detail in the succeeding chapters. Meanwhile, let us assume that the solution of the network is *unique* and has been obtained in terms of the state variables i_{L_j} and v_{C_j}, namely,

$$i_{L_j}(t) = I_j(t) \qquad j = 1, 2, \ldots, k \tag{13-11}$$

$$v_{C_j}(t) = E_j(t) \qquad j = k+1, k+2, \ldots, n \tag{13-12}$$

Observe that if the solution is obtained in terms of the state variables φ_{L_j} and q_{C_j}, we can easily obtain $i_{L_j}(t)$ by a graphical composition between $\varphi_{L_j}(t)$ and the i_{L_j}-φ_{L_j} curve of inductor L_j. Similarly, we can easily obtain $v_{C_j}(t)$ by a graphical composition between $q_{C_j}(t)$ and the v_{C_j}-q_{C_j} curve of capacitor C_j. There is, therefore, no loss of generality to assume that the solution of the normal-form equations is given by Eqs. (13-11) and (13-12).

Observe that these equations represent the solution of the energy-storage subnetwork N_1 only; namely, the inductor currents $i_{L_j}(t)$ and the capacitor voltages $v_{C_j}(t)$. To find the voltage and current solutions of the elements inside the resistive subnetwork N_2, we can apply the current-source substitution theorem and the voltage-source substitution theorem of Chap. 7 and replace the inductors of N_1 by appropriate current sources and capacitors of N_1 by appropriate voltage sources, as shown in Fig. 13-5*c*.† The resulting network is purely *resistive* and is equivalent to the original network. Therefore, we can apply the techniques developed in Part 2 to obtain the solutions inside N_2.

The above observation is a fundamental result. It shows that the solution of the energy-storage subnetwork completely specifies the solution of the resistive subnetwork. This is equivalent to saying that the state variables in the normal-form equation constitute the minimum set of variables from which all other variables in the network can be calculated. Since the order of complexity is equal to the number of state variables in the normal-form equations, we can also conclude that the order of complexity is equal to the minimum number of initial conditions that must be specified in order to specify the solution of the complete network.

13-4 DERIVING NORMAL-FORM EQUATIONS FROM THE CHARACTERISTICS OF THE N-PORT SUBNETWORK

In Chap. 4 the normal-form equations of a dynamic network are obtained by eliminating and substituting variables from the equations of motion. Although this procedure is quite general, it is sometimes very involved. We shall now describe another procedure for obtaining the normal-form equations of the dynamic network shown in Fig. 13-5*b*.[1] This procedure is valid, provided that the number of energy-storage elements is equal to the order of complexity of the network. Hence, we shall assume that the dynamic network contains neither cut sets of inductors and current sources nor loops of capacitors and voltage sources. This assumption is usually satisfied in practice.[2]

13-4-1 NORMAL-FORM EQUATIONS FOR FIRST-ORDER NETWORKS

A first-order network containing one energy-storage element can be represented either by the one-inductor network shown in Fig. 13-6*a* or the one-capacitor network shown in Fig. 13-6*b*.

---

† These theorems are applicable here since we assume that the solutions are unique.

[1] Although this alternate procedure is not necessarily easier for higher-order networks ($n > 2$), it has more conceptual value for our subsequent discussions. Among other things, it reveals the intricate relationship between the energy-storage subnetwork and the resistive subnetwork.

[2] In fact, to be more realistic, we must include the parasitic elements. Hence, if there is a cut set of inductors and current sources, we can always introduce a parasitic capacitor or resistor across one or more inductors in the cut set, thereby eliminating the inductor-current-source-only cut sets. Similarly, a capacitor-voltage-source-only loop can be eliminated by introducing a parasitic inductor or resistor in series with one or more capacitors in the loop.

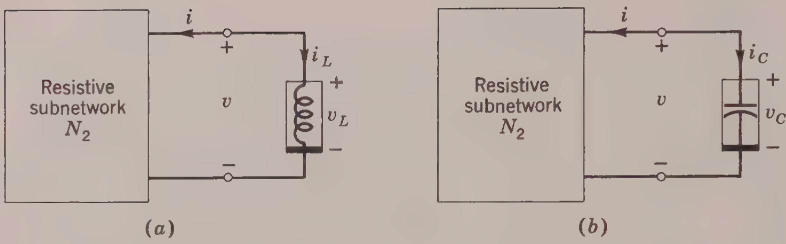


Fig. 13-6. Two first-order networks containing one energy-storage element.

Since N_2 is a resistive network, it has a DP plot at any time t.† Let us denote it by

$$f(i,v,t) = 0 \tag{13-13}$$

At any time t Eq. (13-13) represents one DP plot. This can be derived by the principles described in Part 2.

Consider first the inductor network shown in Fig. 13-6*a*. In this case it is convenient to express Eq. (13-13) in the form

$$v = R(i,t)$$

Since $v = v_L = d\varphi/dt$, we have

$$\frac{d\varphi}{dt} = R(i,t) \tag{13-14}$$

If the inductor is characterized by a flux-controlled curve

$$i_L = i_L(\varphi) \tag{13-15}$$

then since $i = -i_L$, we can substitute Eq. (13-15) for i in Eq. (13-14) to obtain

$$\frac{d\varphi}{dt} = R(-i_L(\varphi),t) \tag{13-16}$$

Equation (13-16) is now in the normal form with the flux linkage φ as the state variable.

EXAMPLE

Consider the first-order network shown in Fig. 13-7*a* containing a neon bulb whose v_1-i_1 curve is shown in Fig. 13-7*b* and a nonlinear inductor described by

---

† See the first fundamental theorem of dc-resistive networks in Sec. 5-6.

$$i_L = \tanh \varphi \tag{13-17}$$

To obtain the normal-form equation, we first redraw the given network into a one-port resistive subnetwork N_2 terminated by the inductor as shown in Fig. 13-7*c*. Our next task is to derive the DP plot across terminals *a-b* of this one-port network. If we denote the current-controlled v_1-i_1 curve of the neon bulb by

$$v_1 = f_1(i_1) \tag{13-18}$$

then by KVL we obtain

$$v = (20)i + f_1(i) + (i - 2 \sin t)(10) - 5$$

where $i = i_1$ is expressed in milliamperes. This equation can be rewritten in the form

$$v = R(i,t) = f_1(i) + 30i - 5 - 20 \sin t \tag{13-19}$$

From this we obtain the normal-form equation upon substituting Eqs. (13-19) and (13-17) into Eq. (13-16); thus

$$\frac{d\varphi}{dt} = f_1(-\tanh \varphi) - 30 \tanh \varphi - 5 - 20 \sin t \tag{13-20}$$

The term $f_1(-\tanh \varphi)$ represents a function of φ obtained by a graphical composition between $f_1(i)$ and $i = -\tanh \varphi$.

If the inductor is characterized by a current-controlled curve

$$\varphi = \Phi(i_L) \tag{13-21}$$

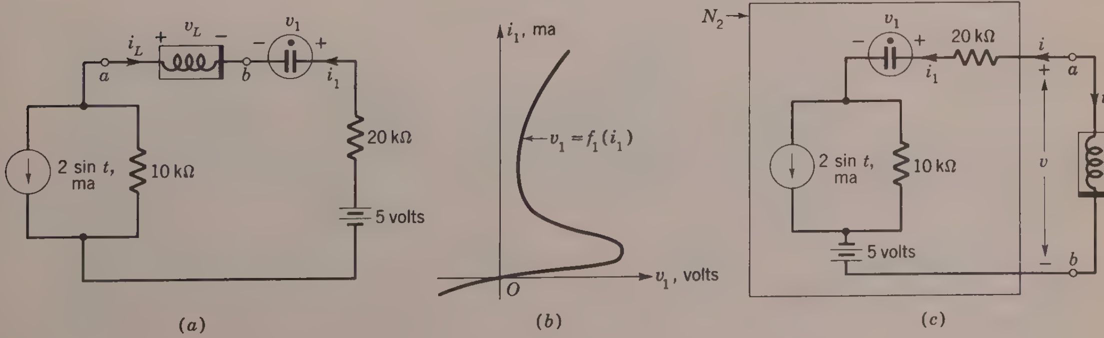


Fig. 13-7. Example of a first-order network containing a neon bulb and a nonlinear inductor.

then by applying the chain rule to Eq. (13-14), we can obtain the normal-form equation in terms of the state variable i_L, namely,

$$\frac{d\varphi}{di_L}\frac{di_L}{dt} = R(-i_L,t)$$

or

$$\frac{di_L}{dt} = \frac{R(-i_L,t)}{L(i_L)} \tag{13-22}$$

where $L(i_L)$ denotes $d\varphi/di_L$, the *incremental inductance.* In the special case where the inductor is linear, $\varphi = Li_L$ and Eq. (13-22) reduces to

$$\frac{di_L}{dt} = \frac{R(-i_L,t)}{L} \tag{13-23}$$

We can derive the normal-form equation for the capacitor network shown in Fig. 13-6*b* by a dual procedure. We first derive the DP plot across terminals *a-b* in the form

$$i = G(v,t)$$

If the capacitor is characterized by a charge-controlled curve $v_C = v_C(q)$, then the normal-form equation takes the form

$$\frac{dq}{dt} = -G(v_C(q),t) \tag{13-24}$$

where the capacitor charge q is the state variable.

On the other hand, if the capacitor is characterized by a voltage-controlled curve $q = Q(v_C)$, then the normal-form equation takes the form

$$\frac{dv_C}{dt} = -\frac{G(v_C,t)}{C(v_C)} \tag{13-25}$$

where the capacitor voltage v_C is the state variable and where $C(v_C)$ denotes dq/dv_C, the *incremental capacitance.* In the special case where the capacitor is linear, $q = Cv_C$ and Eq. (13-25) reduces to

$$\frac{dv_C}{dt} = -\frac{G(v_C,t)}{C} \tag{13-26}$$

Observe that in all cases, the normal-form equation of a first-order network is always of the form

$$\frac{dx}{dt} = f(x,t) \tag{13-27}$$

The state variable x is either φ or i_L for an inductor network and is either q or v_C for a capacitor network. Observe that in all cases, the function $f(x,t)$ in Eq. (13-27) is obtained by deriving the v-i characteristics of the one-port resistive subnetwork N_2.

In the special case where the network is autonomous,[1] Eq. (13-27) reduces to

$$\frac{dx}{dt} = f(x) \tag{13-28}$$

In this case the function $f(x)$ consists of only one curve which can be easily derived from the DP plot of N_2.

13-4-2 NORMAL-FORM EQUATIONS FOR SECOND-ORDER NETWORKS

A second-order network containing two energy-storage elements may assume one of the three possible forms shown in Fig. 13-8*a* to *c*. To obtain the normal-form equations, the first step is to derive the relationship between the variables i_1, i_2, v_1, and v_2 of the two-port resistive subnetwork N_2.

Consider first the two-inductor network shown in Fig. 13-8*a*. Let us assume that the characteristics of N_2 have been derived and expressed in the following form[2]:

$$v_1 = R_1(i_1,i_2,t) \tag{13-29}$$

$$v_2 = R_2(i_1,i_2,t) \tag{13-30}$$

If the inductors are characterized by $i_{L_1} = i_{L_1}(\varphi_1)$ and $i_{L_2} = i_{L_2}(\varphi_2)$, then since $v_{L_1} = v_1 = d\varphi_1/dt$, $v_{L_2} = v_2 = d\varphi_2/dt$ and $i_1 = -i_{L_1}$, $i_2 = -i_{L_2}$, the normal-form equations in terms of the state variables φ_1 and φ_2 are given by

$$\frac{d\varphi_1}{dt} = R_1(-i_{L_1}(\varphi_1), -i_{L_2}(\varphi_2), t) \tag{13-31}$$

$$\frac{d\varphi_2}{dt} = R_2(-i_{L_1}(\varphi_1), -i_{L_2}(\varphi_2), t) \tag{13-32}$$

[1] Recall that a dynamic network is autonomous if it does not contain ac sources.

[2] Review Sec. 3-2-1 on the six common forms of representation.

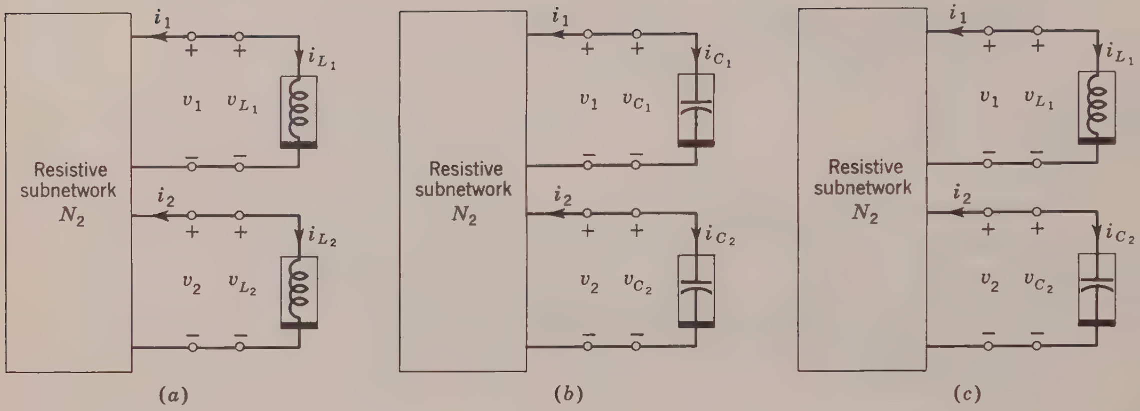


Fig. 13-8. Three possible second-order networks containing two energy-storage elements.

On the other hand, if the inductors are characterized by $\varphi_1 = \Phi_1(i_{L_1})$ and $\varphi_2 = \Phi_2(i_{L_2})$, then by applying the chain rule, we obtain the following normal-form equations in terms of the state variables i_{L_1} and i_{L_2}:

$$\frac{di_{L_1}}{dt} = \frac{R_1(-i_{L_1},-i_{L_2},t)}{L_1(i_{L_1})} \tag{13-33}$$

$$\frac{di_{L_2}}{dt} = \frac{R_2(-i_{L_1},-i_{L_2},t)}{L_2(i_{L_2})} \tag{13-34}$$

where $L_1(i_{L_1})$ denotes $d\varphi_1/di_{L_1}$, and $L_2(i_{L_2})$ denotes $d\varphi_2/di_{L_2}$. In the special case where both inductors are linear, $\varphi_1 = L_1 i_{L_1}$ and $\varphi_2 = L_2 i_{L_2}$, Eqs. (13-33) and (13-34) reduce to

$$\frac{di_{L_1}}{dt} = \frac{R_1(-i_{L_1},-i_{L_2},t)}{L_1} \tag{13-35}$$

$$\frac{di_{L_2}}{dt} = \frac{R_2(-i_{L_1},-i_{L_2},t)}{L_2} \tag{13-36}$$

The normal-form equation for the two-capacitor network can be obtained by dual procedures. In this case we derive and express the characteristics of the two-port resistive network N_2 in the following form:

$$i_1 = G_1(v_1,v_2,t) \tag{13-37}$$

$$i_2 = G_2(v_1,v_2,t) \tag{13-38}$$

If the capacitors are characterized by $v_{C_1} = v_{C_1}(q_1)$ and $v_{C_2} = v_{C_2}(q_2)$, then the normal-form equations in terms of the state variables q_1 and q_2 are given by

$$\frac{dq_1}{dt} = -G_1(v_{C_1}(q_1), v_{C_2}(q_2), t) \tag{13-39}$$

$$\frac{dq_2}{dt} = -G_2(v_{C_1}(q_1), v_{C_2}(q_2), t) \tag{13-40}$$

If the capacitors are characterized by $q_1 = Q_1(v_{C_1})$ and $q_2 = Q_2(v_{C_2})$, then the normal-form equations in terms of the state variables v_{C_1} and v_{C_2} are given by

$$\frac{dv_{C_1}}{dt} = \frac{-G_1(v_{C_1}, v_{C_2}, t)}{C_1(v_{C_1})} \tag{13-41}$$

$$\frac{dv_{C_2}}{dt} = \frac{-G_2(v_{C_1}, v_{C_2}, t)}{C_2(v_{C_2})} \tag{13-42}$$

where $C_1(v_{C_1})$ denotes dq_1/dv_{C_1}, and $C_2(v_{C_2})$ denotes dq_2/dv_{C_2}.

In the special case where both capacitors are linear with capacitances C_1 and C_2, respectively, Eqs. (13-41) and (13-42) reduce to

$$\frac{dv_{C_1}}{dt} = \frac{-G_1(v_{C_1}, v_{C_2}, t)}{C_1} \tag{13-43}$$

$$\frac{dv_{C_2}}{dt} = \frac{-G_2(v_{C_1}, v_{C_2}, t)}{C_2} \tag{13-44}$$

For the inductor-capacitor network shown in Fig. 13-8*c*, it is necessary to derive the characteristics of N_2 in the following hybrid form:

$$v_1 = H_1(i_1, v_2, t) \tag{13-45}$$

$$i_2 = H_2(i_1, v_2, t) \tag{13-46}$$

If the inductor and the capacitor are characterized, respectively, by $i_{L_1} = i_{L_1}(\varphi_1)$ and $v_{C_2} = v_{C_2}(q_2)$, then the normal-form equations in terms of the state variables φ_1 and q_2 are given by

$$\frac{d\varphi_1}{dt} = H_1(-i_{L_1}(\varphi_1), v_{C_2}(q_2), t) \tag{13-47}$$

$$\frac{dq_2}{dt} = -H_2(-i_{L_1}(\varphi_1), v_{C_2}(q_2), t) \tag{13-48}$$

If the inductor and capacitor are characterized, respectively, by $\varphi_1 = \Phi_1(i_{L_1})$ and $q_2 = Q_2(v_{C_2})$, then the normal-form equations in terms of the state variables i_{L_1} and v_{C_2} are given by

$$\frac{di_{L_1}}{dt} = \frac{H_1(-i_{L_1}, v_{C_2}, t)}{L_1(i_{L_1})} \tag{13-49}$$

$$\frac{dv_{C_2}}{dt} = \frac{-H_2(-i_{L_1}, v_{C_2}, t)}{C_2(v_{C_2})} \tag{13-50}$$

where $L_1(i_{L_1})$ denotes $d\varphi_1/di_{L_1}$, and $C_2(v_{C_2})$ denotes dq_2/dv_{C_2}.

In the special case where the inductor is linear with inductance L_1 and the capacitor is linear with capacitance C_2, Eqs. (13-49) and (13-50) reduce to

$$\frac{di_{L_1}}{dt} = \frac{H_1(-i_{L_1}, v_{C_2}, t)}{L_1} \tag{13-51}$$

$$\frac{dv_{C_2}}{dt} = \frac{-H_2(-i_{L_1}, v_{C_2}, t)}{C_2} \tag{13-52}$$

Observe that in all cases, the normal-form equations of a second-order network are always of the form

$$\frac{dx_1}{dt} = f_1(x_1, x_2, t) \tag{13-53}$$

$$\frac{dx_2}{dt} = f_2(x_1, x_2, t) \tag{13-54}$$

In the special case where the network is autonomous, Eqs. (13-53) and (13-54) reduce to

$$\frac{dx_1}{dt} = f_1(x_1, x_2) \tag{13-55}$$

$$\frac{dx_2}{dt} = f_2(x_1, x_2) \tag{13-56}$$

Observe that in each case, the functions f_1 and f_2 are easily obtained from the relationships characterizing the two-port resistive subnetwork N_2.

EXAMPLE

To illustrate the above procedure, let us derive the normal-form equation for the network shown in Fig. 4-15 of Chap. 4. First, let us redraw this network in the form of a two-port resistive network N_2 terminated in an inductor and a capacitor, as shown in Fig. 13-9. The characteristics of the elements are given as before.

Three-terminal resistor

$$i_1 = 1 + v_1 + 3v_1 i_2^3 - 4i_2^5 \tag{13-57}$$

$$v_2 = 4 - i_2 v_1 - 2i_2^2 v_1^5 + v_1^3 \tag{13-58}$$

Capacitor

$$q_3 = 2 - 3v_3^3 + 5v_3^5 \tag{13-59}$$

Inductor

$$\varphi_4 = 1 + 2i_4 - 3i_4^2 + i_4^3 \tag{13-60}$$

In accordance with Eqs. (13-45) and (13-46) our first step is to derive the relationship characterizing N_2 in the following form:

$$v_a = H_1(i_a, v_b, t)$$

$$i_b = H_2(i_a, v_b, t)$$

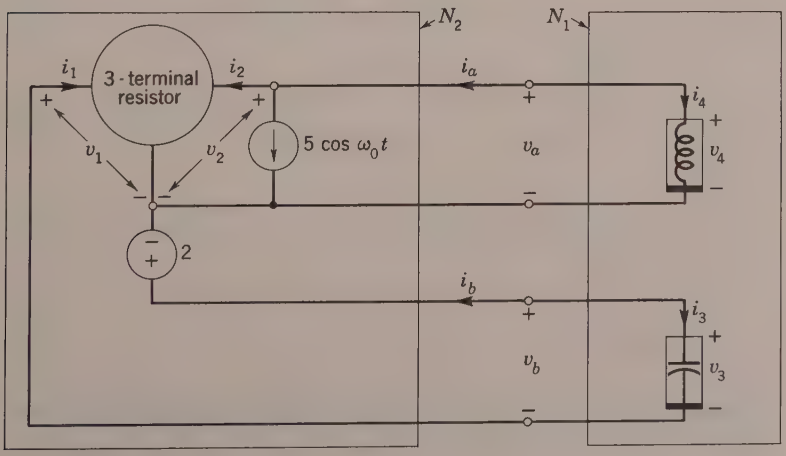


Fig. 13-9. To derive the normal-form equations, the network of Fig. 4-15 is redrawn into a resistive subnetwork and an energy-storage subnetwork.

From Fig. 13-9 we obtain

$$v_1 = 2 - v_b \tag{13-61}$$

$$i_1 = -i_b \tag{13-62}$$

$$v_2 = v_a \tag{13-63}$$

$$i_2 = -5 \cos \omega_0 t + i_a \tag{13-64}$$

Substituting Eqs. (13-61) to (13-64) for i_1, i_2, v_1, and v_2 in Eqs. (13-57) and (13-58), we obtain

$$\begin{aligned} v_a = 4 &- (-5 \cos \omega_0 t + i_a)(2 - v_b) \\ &- 2(-5 \cos \omega_0 t + i_a)^2(2 - v_b)^5 + (2 - v_b)^3 \end{aligned} \tag{13-65}$$

$$\begin{aligned} i_b = -1 &- (2 - v_b) - 3(2 - v_b)(-5 \cos \omega_0 t + i_a)^3 \\ &+ 4(-5 \cos \omega_0 t + i_a)^5 \end{aligned} \tag{13-66}$$

From Eqs. (13-59) and (13-60) we obtain

$$\frac{dq_3}{dv_3} \equiv C_3(v_3) = -9v_3^2 + 25v_3^4 \tag{13-67}$$

$$\frac{d\varphi_4}{di_4} \equiv L_4(i_4) = 2 - 6i_4 + 3i_4^2 \tag{13-68}$$

If we substitute Eqs. (13-65) to (13-68) into the corresponding terms in Eqs. (13-49) and (13-50), we obtain the following normal-form equations:

$$\frac{di_4}{dt} = \frac{\begin{aligned} &4 - (-5 \cos \omega_0 t - i_4)(2 - v_3) \\ &\qquad - 2(-5 \cos \omega_0 t - i_4)^2(2 - v_3)^5 - (2 - v_3)^3 \end{aligned}}{2 - 6i_4 + 3i_4^2}$$

$$\frac{dv_3}{dt} = \frac{\begin{aligned} &1 + (2 - v_3) + 3(2 - v_3)(-5 \cos \omega_0 t - i_4)^3 \\ &\qquad - 4(-5 \cos \omega_0 t - i_4)^5 \end{aligned}}{-9v_3^2 + 25v_3^4}$$

Observe that the above equations are identical with Eqs. (4-49*a*) and (4-49*b*) derived earlier in Chap. 4.

13-4-3 NORMAL-FORM EQUATIONS FOR *n*TH-ORDER NETWORKS

An *n*th-order network containing *n* energy-storage elements may assume the general form shown in Fig. 13-5*b*, where $0 \leq k \leq n$.

When $k = n$, the network becomes an n-inductor network, and when $k = 0$, the network becomes an n-capacitor network.

To obtain the normal-form equations, we must derive the characteristics of the n-port resistive network N_2 in the *hybrid form*.

$$\begin{aligned}
v_1 &= H_1(i_1,i_2,\ldots,i_k,v_{k+1},v_{k+2},\ldots,v_n,t)\\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\\
v_k &= H_k(i_1,i_2,\ldots,i_k,v_{k+1},v_{k+2},\ldots,v_n,t)\\
i_{k+1} &= H_{k+1}(i_1,i_2,\ldots,i_k,v_{k+1},v_{k+2},\ldots,v_n,t)\\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\\
i_n &= H_n(i_1,i_2,\ldots,i_k,v_{k+1},v_{k+2},\ldots,v_n,t)
\end{aligned} \tag{13-69}$$

If the inductors and capacitors are characterized, respectively, by

$$i_{L_j} = i_{L_j}(\varphi_j) \qquad j = 1, 2, \ldots, k \tag{13-70}$$

$$v_{C_j} = v_{C_j}(q_j) \qquad j = k+1, k+2, \ldots, n \tag{13-71}$$

then the normal-form equations in terms of the state variables φ_1, $\varphi_2, \ldots, \varphi_k, q_{k+1}, q_{k+2}, \ldots, q_n$ are given by

$$\begin{aligned}
\frac{d\varphi_1}{dt} &= H_1(-i_{L_1}(\varphi_1),-i_{L_2}(\varphi_2),\ldots,\\
&\qquad -i_{L_k}(\varphi_k),v_{C_{k+1}}(q_{k+1}),\ldots,v_{C_n}(q_n),t)\\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\\
\frac{d\varphi_k}{dt} &= H_k(-i_{L_1}(\varphi_1),-i_{L_2}(\varphi_2),\ldots,\\
&\qquad -i_{L_k}(\varphi_k),v_{C_{k+1}}(q_{k+1}),\ldots,v_{C_n}(q_n),t)\\
\frac{dq_{k+1}}{dt} &= -H_{k+1}(-i_{L_1}(\varphi_1),-i_{L_2}(\varphi_2),\ldots,\\
&\qquad -i_{L_k}(\varphi_k),v_{C_{k+1}}(q_{k+1}),\ldots,v_{C_n}(q_n),t)\\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\\
\frac{dq_n}{dt} &= -H_n(-i_{L_1}(\varphi_1),-i_{L_2}(\varphi_2),\ldots,\\
&\qquad -i_{L_k}(\varphi_k),v_{C_{k+1}}(q_{k+1}),\ldots,v_{C_n}(q_n),t)
\end{aligned} \tag{13-72}$$

Similar expressions can be derived for the case where the inductors are current-controlled and capacitors are voltage-controlled or for the case where all elements are linear. In all cases the normal-form equations of an nth-order network are always of the form

$$\begin{aligned}\frac{dx_1}{dt} &= f_1(x_1,x_2, \ldots ,x_n,t)\\ \frac{dx_2}{dt} &= f_2(x_1,x_2, \ldots ,x_n,t)\\ &\cdots\cdots\cdots\cdots\cdots\\ \frac{dx_n}{dt} &= f_n(x_1,x_2, \ldots ,x_n,t)\end{aligned} \tag{13-73}$$

In the special case where the network is autonomous, the normal-form equations reduce to

$$\begin{aligned}\frac{dx_1}{dt} &= f_1(x_1,x_2, \ldots ,x_n)\\ \frac{dx_2}{dt} &= f_2(x_1,x_2, \ldots ,x_n)\\ &\cdots\cdots\cdots\cdots\cdots\\ \frac{dx_n}{dt} &= f_n(x_1,x_2, \ldots ,x_n)\end{aligned} \tag{13-74}$$

In either case, the right-hand side of these equations can be easily obtained from the relationships characterizing the n-port resistive network N_2.

Exercise 1. List all possible combinations of state variables in the normal-form equations for a network with three energy-storage elements.

Exercise 2: Find the general expressions for the normal-form equations of an nth-order network corresponding to the following cases: (*a*) All inductors are current-controlled, and all capacitors are voltage-controlled. (*b*) All energy-storage elements are linear.

13-5 TRAJECTORY OF SOLUTIONS IN THE STATE SPACE

Consider an arbitrary nth-order network whose equations of motion have been obtained in the normal form. One of the most fundamental problems in dynamic network theory consists of determining the solution of this system of n first-order nonlinear differential equations subject to a prescribed set of initial conditions; namely, $x_1 = x_{1_0}$, $x_2 = x_{2_0}$, . . . , and $x_n = x_{n_0}$. The solution consists of n time functions $x_1 = X_1(t)$, $x_2 = X_2(t)$, . . . , $x_n = X_n(t)$

such that upon substituting these into Eq. (13-73), we obtain the following two sets of identities:[1]

$$\begin{aligned}
\frac{dX_1(t)}{dt} &\equiv f_1(X_1(t),X_2(t),\ldots,X_n(t),t)\\
\frac{dX_2(t)}{dt} &\equiv f_2(X_1(t),X_2(t),\ldots,X_n(t),t)\\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\\
\frac{dX_n(t)}{dt} &\equiv f_n(X_1(t),X_2(t),\ldots,X_n(t),t)
\end{aligned} \tag{13-75}$$

$$\begin{aligned}
X_1(t_0) &\equiv x_{1_0}\\
X_2(t_0) &\equiv x_{2_0}\\
&\cdots\cdots\\
X_n(t_0) &\equiv x_{n_0}
\end{aligned} \tag{13-76}$$

With the present state of the art it is very difficult and often impossible to determine the solutions $X_1(t)$, $X_2(t)$, . . . , $X_n(t)$ in exact analytical form. Fortunately, there are many practical problems where it is not necessary to know the exact analytical solution. Instead, only certain *qualitative* aspects of the solution are desired. For example, in the design of many practical electronic circuits it is necessary to know the maximum voltage across some nonlinear device so that its maximum voltage rating can be specified. In this case we are interested in only one property of the solution, namely, its maximum value. In other problems we may be interested primarily in knowing whether the solution $X_j(t)$ eventually settles down asymptotically to a constant value or eventually oscillates periodically. We shall refer to this property of the solution as the *steady-state behavior* of the dynamic network. In order to investigate the qualitative aspects of the solution, it is necessary to understand a number of new basic concepts to be introduced in the following sections. These concepts are formulated in terms of the trajectory of solutions described below.

There are several ways to interpret the solutions $x_1 = X_1(t)$, $x_2 = X_2(t)$, . . . , $x_n = X_n(t)$ of a dynamic network. The most common method is to interpret each solution $x_j = X_j(t)$ as a curve. In this case n separate curves are needed to completely specify the solutions. Another method is to represent the n state variables $x_1, x_2, \ldots, x_n$ as the coordinate axes of an n-dimensional space, henceforth called the *state space*. At any time $t = t_j$, the solutions

[1] Notice that we have chosen the capital letter X_j to denote the solution waveform, as a function of time, of the corresponding state variable x_j.

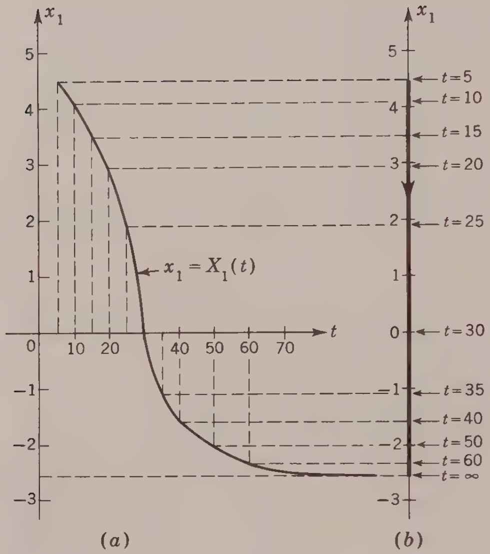


Fig. 13-10. Trajectory of a monotonic time function.

$x_1 = X_1(t_j)$, $x_2 = X_2(t_j)$, . . . , $x_n = X_n(t_j)$ can be interpreted as the coordinates of a point $P(t_j)$ in the state space. If we imagine that the above interpretation is applied to all times $t \geq t_0$, the resulting points can be connected to form a smooth curve in the state space. This curve is called the *trajectory of solutions,* or simply, the *trajectory.*[1] Observe that the parameter t associated with each point along the trajectory must be specified in order to recover the solutions. A trajectory is incomplete unless this parametrization is included.

It is common practice to refer to any point $(x_1,x_2, \ldots ,x_n)$ in the state space as a *state* of the dynamic network. For example, the initial condition $(x_{1_0},x_{2_0}, \ldots ,x_{n_0})$ is usually called the *initial state.* A trajectory, therefore, always starts from an initial state corresponding to $t = t_0$.

To illustrate the concept of the trajectory of solutions, let us consider some specific cases.

Case 1: Trajectory of first-order networks When $n = 1$, the solution consists of one time function $x_1 = X_1(t)$. In this case the state space is one-dimensional, and the trajectory is simply a segment of the coordinate axis x_1. For example, if $X_1(t)$ is as shown in Fig. 13-10*a*, then the trajectory is obtained by projecting the coordinates of each point of $X_1(t)$ into a vertically aligned axis with the corresponding time interval labeled on it as shown in Fig. 13-10*b*. It should be clear that if the time function $X_1(t)$ is not monotonic, then some

---

[1] This term is borrowed from classical mechanics, where it is used to denote the path taken by a particle in motion.

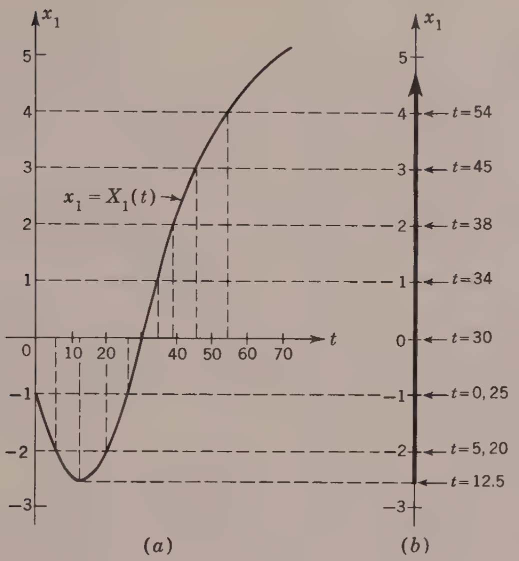


Fig. 13-11. Trajectory of a nonmonotonic time function.

points along the trajectory will contain more than one time parameter. For example, if $X_1(t)$ is as shown in Fig. 13-11*a*, then its trajectory is readily seen to be given by Fig. 13-11*b*. Observe that in either case we can reverse the procedure to recover the time function $X_1(t)$ from its trajectory.

Case 2: Trajectory of second-order networks When $n=2$, the solution consists of two time functions $x_1 = X_1(t)$ and $x_2 = X_2(t)$. In this case the state space is two-dimensional, and the trajectory consists of a curve in the x_1-x_2 plane, otherwise known as the *phase plane.*[1] For example, if $X_1(t)$ and $X_2(t)$ are as shown in Fig. 13-12*a*, then the corresponding trajectory is given by Fig. 13-12*b*. This trajectory is called an *open trajectory* because each point on it corresponds to only one instant of time. It is possible to have a trajectory where each point on it corresponds to more than one instant of time. For example, if $X_1(t)$ and $X_2(t)$ are as shown in Fig. 13-13*a*, then the corresponding trajectory is shown in Fig. 13-13*b*. Observe that this trajectory forms a closed loop, and hence it is called a *closed trajectory.* Observe also that for the simple waveforms shown in Fig. 13-13*a*, the corresponding trajectory could have been obtained analytically by eliminating the variable t; namely,

$$\frac{x_1^2}{4} + \frac{x_2^2}{9} = \left(\sin \frac{\pi}{2} t\right)^2 + \left(\cos \frac{\pi}{2} t\right)^2 = 1$$

---

[1] When $n = 2$, the state space is usually called the *phase plane* for historical reasons. In the past, one of the two state variables was usually called the *phase,* and for $n = 2$, the word *plane* is preferred over the word *space.*

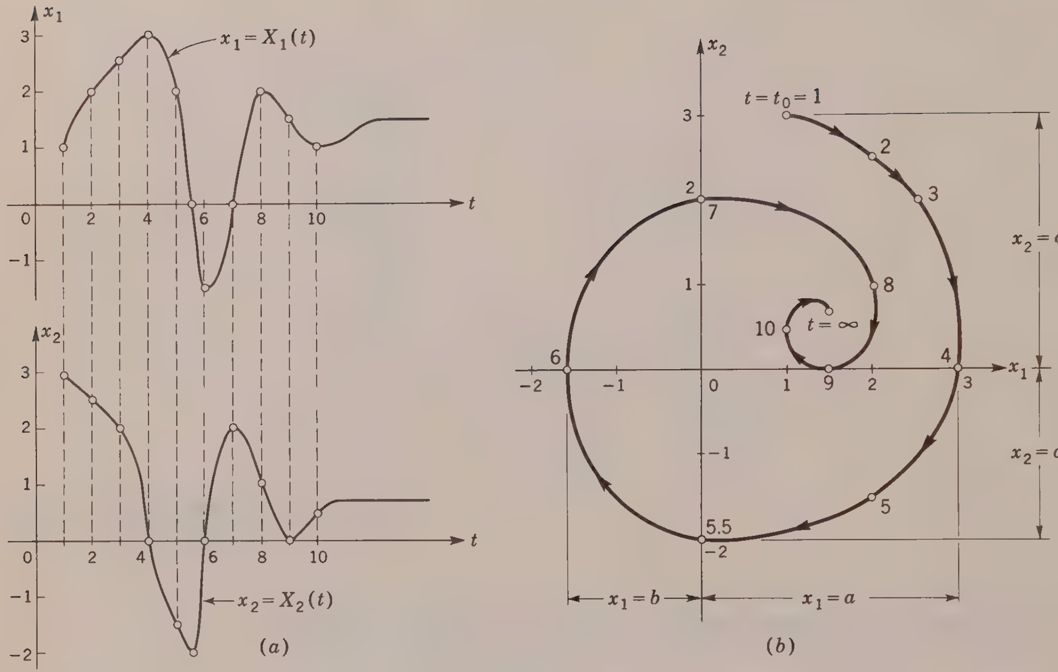


Fig. 13-12. Example of an open trajectory of a second-order network.

Finally, it is also possible to obtain a trajectory which eventually approaches a closed loop known as a *limit cycle.*[1] For example, if $X_1(t)$ and $X_2(t)$ are as shown in Fig. 13-14*a*, then the corresponding trajectory is given by Fig. 13-14*b*. Observe that the trajectory will eventually approach a limit cycle. Clearly, there are many trajectories other than those described above. For example, if $X_1(t)$ and $X_2(t)$ are as shown in Fig. 13-15*a*, then the corresponding trajectory is as shown in Fig. 13-15*b*. Observe that in all cases, the trajectories of second-order networks are simply obtained by a graphical elimination procedure. Observe also that if we apply $X_1(t)$ and $X_2(t)$ to the horizontal and vertical terminals of an oscilloscope, the resulting Lissajous figure is exactly the trajectory corresponding to $X_1(t)$ and $X_2(t)$.

From the above examples it is clear that if we can obtain the trajectory, then some qualitative information about the network will be known even if the time parameter is not labeled on it. If we know that the trajectory is open and does not approach ∞, then we know that $X_1(t)$ and $X_2(t)$ must eventually approach a constant value and are therefore bounded. The trajectory of Fig. 13-12*b* shows that $b < x_1 < a$ and $d < x_2 < c$. This information might be used to determine whether the rating of a device is ex-

---

[1] Theoretically speaking, the trajectory will approach the closed loop only in the limit when $t \to \infty$, hence the term limit cycle.

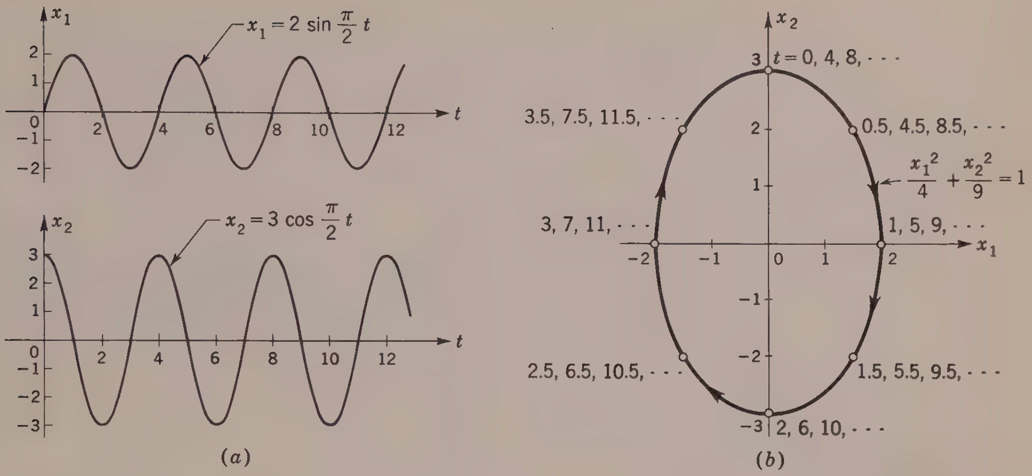
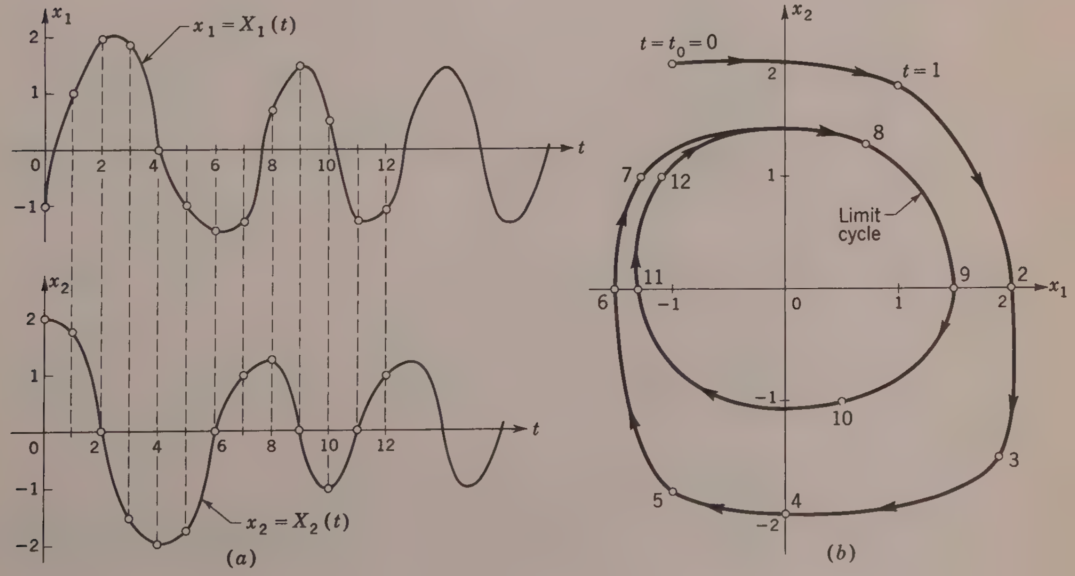


Fig. 13-13. Example of a closed trajectory of a second-order network.

ceeded or not. On the other hand, if we know that the trajectory eventually approaches a limit cycle, then we know that $X_1(t)$ and $X_2(t)$ would eventually approach some periodic waveforms, in which case we can conclude that the network is oscillatory.

Case 3: Trajectory of nth-order networks When $n > 2$, the solution consists of n time functions $x_1 = X_1(t)$, $x_2 = X_2(t)$, . . . , $x_n = X_n(t)$. In this case the state space is n-dimensional, and the trajectory

Fig. 13-14. Example of a trajectory which approaches a limit cycle.



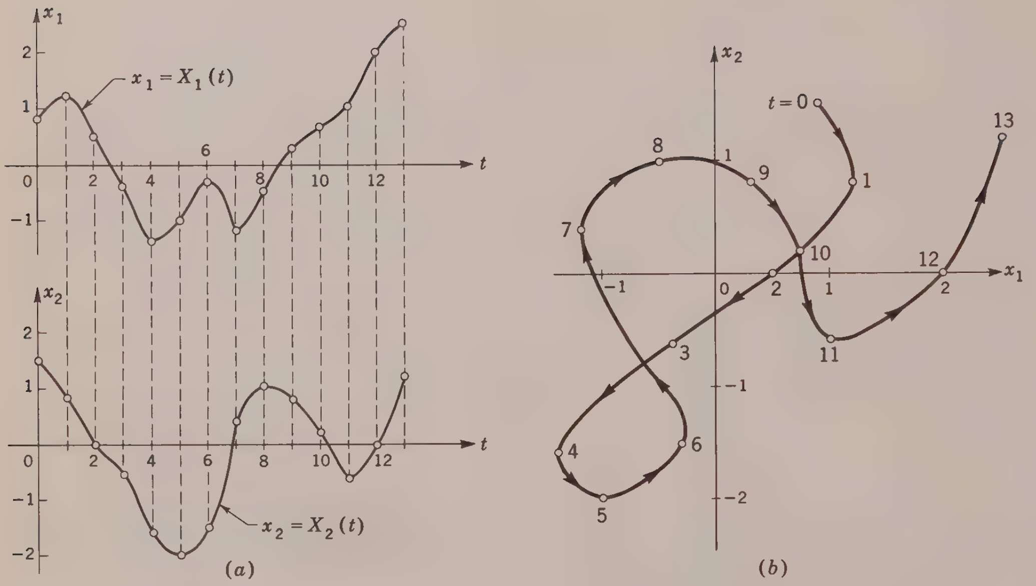


Fig. 13-15. Example of an arbitrary trajectory of a second-order network.

consists of a curve in the x_1-x_2-$\cdots$-x_n space. For example, the trajectory corresponding to the three time functions $X_1(t)$, $X_2(t)$, and $X_3(t)$ shown in Fig. 13-16*a* is readily seen to be given by Fig. 13-16*b*. It is important to observe that this trajectory is a *curve* and *not* a surface in the three-dimensional state space.

When $n > 3$, it is no longer possible to draw the trajectory, but nevertheless it is still useful to visualize the trajectory of an nth-order network as a curve (not a surface) in an n-dimensional state space. As a matter of fact, the above concept is extremely important, and has found wide application in many diverse fields of study such as control systems, biological systems, and even economic systems.

Exercise 1: Sketch the trajectories corresponding to the following solutions of some first-order networks: (*a*) $v_C(t) = 10e^{-t} - 5$, $t \geq 0$. (*b*) $i_L(t) = 100e^{-0.1t}$, $0 < t < 10$; $i_L(t) = 200e^{-0.2(t-10)}$, $10 \leq t < \infty$.

Exercise 2: Sketch the trajectories corresponding to the following solutions of some second-order networks: (*a*) $q_1(t) = 10 \sin t$, $q_2(t) = 20 \cos t$, $t \geq 0$. (*b*) $q_1(t) = 10e^{-t} \sin t$, $q_2(t) = 20e^{-t} \cos t$, $t \geq 0$. (*c*) $q_1(t) = 10e^{t} \sin t$, $q_2(t) = 20e^{t} \cos t$, $t \geq 0$.

Exercise 3: Give a simple test for determining whether the trajectory corresponding to $X_1(t)$ and $X_2(t)$ will or will not intersect itself.

Exercise 4: If the following statements are true, prove them. If not true, give counter-examples and state the conditions under which they are true. (*a*) The trajectory of a first-order network has a uniform time scale if, and only if, the corresponding solution $X(t)$ is a piecewise-linear function of time. (*b*) The trajectory of a second-order network corresponding to $X_1(t)$ and $X_2(t)$ is piecewise-linear if, and only if, both $X_1(t)$ and $X_2(t)$ are piecewise-linear.

13-6 EQUILIBRIUM STATES OF AUTONOMOUS NETWORKS

Intuitively, a physical system is said to be in *equilibrium* if its *motion* does not vary with time. By motion we mean the solutions to the normal-form equations of the system. In a dynamic network, the motion corresponds to the solutions $x_1 = X_1(t), x_2 = X_2(t), \ldots,$ $x_n = X_n(t)$ of Eq. (13-73) or Eq. (13-74), where $x_1, x_2, \ldots, x_n$ are

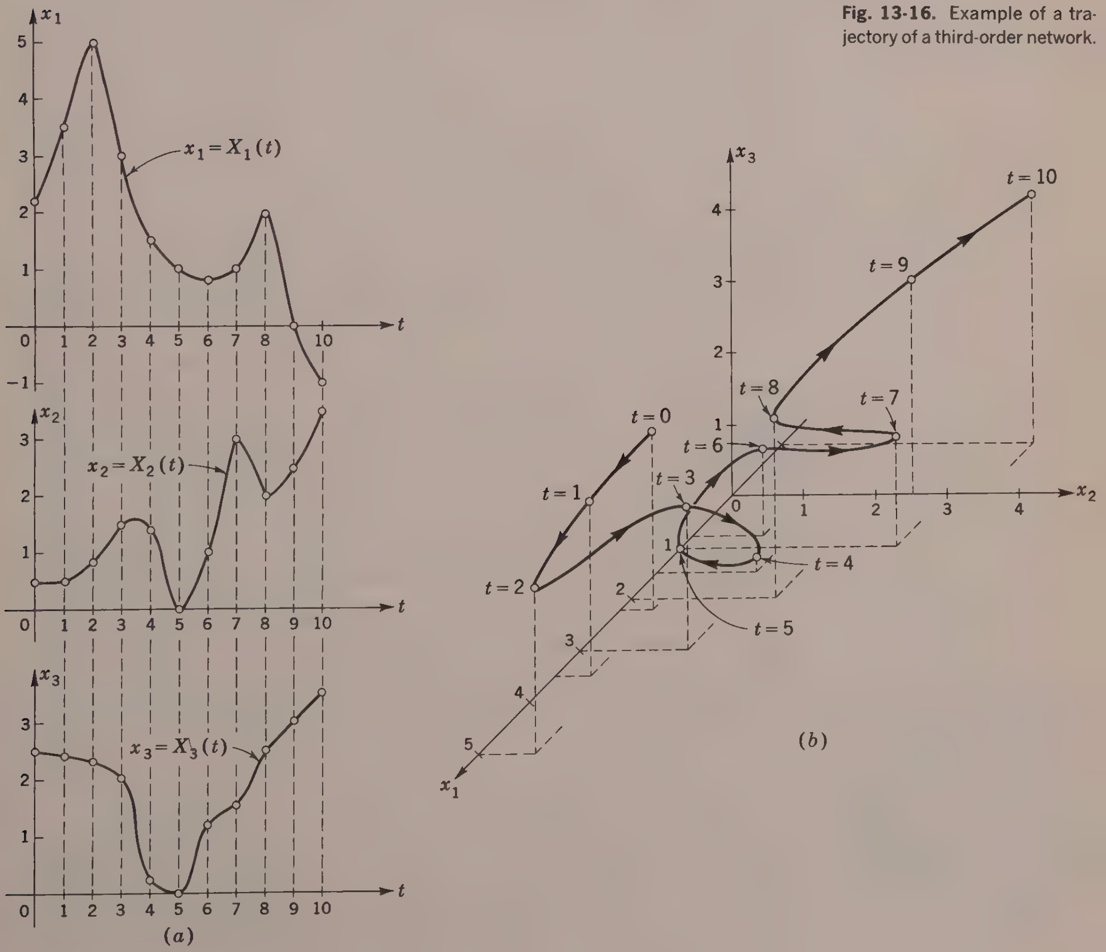


Fig. 13-16. Example of a trajectory of a third-order network.

the state variables. Hence, the network is in equilibrium if, and only if,

$$\begin{aligned} \frac{dx_1}{dt} &\equiv 0 \\ \frac{dx_2}{dt} &\equiv 0 \\ &\cdots\cdots \\ \frac{dx_n}{dt} &\equiv 0 \end{aligned} \tag{13-77}$$

simultaneously. These identities imply that the solutions are constants; namely,

$$\begin{aligned} x_1 &= k_1 \\ x_2 &= k_2 \\ &\cdots\cdots \\ x_n &= k_n \end{aligned} \tag{13-78}$$

This equilibrium solution $(k_1,k_2, \ldots ,k_n)$ represents a point in the state space and is henceforth called an *equilibrium state* of the network.[1]

An examination of Eq. (13-73) of a nonautonomous network reveals the impossibility of obtaining a constant solution in the form of Eq. (13-78). This is because the terms on the right-hand side of Eq. (13-73) contain the parameter t, and therefore the solution must also be functions of t; otherwise, the two sides of this equation cannot be identical. This observation shows that only autonomous networks can be in equilibrium. Accordingly, only autonomous networks will be considered in the following sections.

It is desirable to know the equilibrium state of a network because it represents the steady-state behavior. To determine the equilibrium state, we must impose the constraint given by Eq. (13-77) on the normal-form equations of an autonomous network (Eq. 13-75); namely,

$$\begin{aligned} f_1(x_1,x_2, \ldots ,x_n) &= 0 \\ f_2(x_1,x_2, \ldots ,x_n) &= 0 \\ &\cdots\cdots\cdots\cdots \\ f_n(x_1,x_2, \ldots ,x_n) &= 0 \end{aligned} \tag{13-79}$$

---

[1] In the theory of differential equations the equilibrium states corresponding to a system of autonomous differential equations in the normal form are called *singular points.* We shall use the term equilibrium state in this book because it conveys more physical meaning.

The solution of this system of equations is the equilibrium state. Since Eq. (13-79) is nonlinear, it may have more than one solution; each one qualifies as an equilibrium state. Therefore, an autonomous nonlinear network generally has more than one equilibrium state. We shall summarize the above observations as a theorem.

IMPLICIT EQUILIBRIUM STATE THEOREM

The equilibrium states of any autonomous network can be determined by solving the system of nonlinear functional equations obtained by equating the right-hand side of the normal-form equations of the network to zero.

Before we can apply the implicit equilibrium state theorem, it is necessary to derive the normal-form equation first. For a complicated network, this step is sometimes very involved. Fortunately, the derivation of normal-form equations can be obviated by a proper interpretation of Eq. (13-77). To be specific, let us consider an autonomous network containing k inductors and $(n\text{-}k)$ capacitors, as shown in Fig. 13-17*a*. The first k state variables may represent either inductor flux linkages φ_j or inductor currents i_{L_j}. Similarly, the remaining $(n\text{-}k)$ state variables may represent either capacitor charges q_j or capacitor voltages v_{C_j}. Accordingly, Eq. (13-77) may assume either the form[1]

$$\begin{aligned} \frac{d\varphi_j}{dt} &= 0 \qquad j = 1, 2, \ldots, k \\ \frac{dq_j}{dt} &= 0 \qquad j = (k+1), (k+2), \ldots, n \end{aligned} \tag{13-80}$$

or the form

$$\begin{aligned} \frac{di_{L_j}}{dt} &= 0 \qquad j = 1, 2, \ldots, k \\ \frac{dv_{C_j}}{dt} &= 0 \qquad j = (k+1), (k+2), \ldots, n \end{aligned} \tag{13-81}$$

But

$$\begin{aligned} v_{L_j} &= \frac{d\varphi_j}{dt} = \frac{d\varphi_j}{di_{L_j}} \frac{di_{L_j}}{dt} \\ i_{C_j} &= \frac{dq_j}{dt} = \frac{dq_j}{dv_{C_j}} \frac{dv_{C_j}}{dt} \end{aligned} \tag{13-82}$$

---

[1] On certain occasions it may be desirable to choose a combination of state variables made up of capacitor voltages and charges, as well as inductor currents and flux linkages. In this mixed case, Eq. (13-77) assumes the form of both Eqs. (13-80) and (13-81).

Therefore, in either case Eq. (13-80) or Eq. (13-81) is equivalent to

$$\begin{aligned} v_{L_j} &= 0 \qquad j = 1, 2, \ldots, k \\ i_{C_j} &= 0 \qquad j = (k + 1), (k + 2), \ldots, n \end{aligned} \tag{13-83}$$

We have shown that in the case of an autonomous network, the constraint imposed by Eq. (13-77) is equivalent to the constraint given by Eq. (13-83).

But Eq. (13-83) can be interpreted simply as replacing each inductor by a short circuit and each capacitor by an open circuit, as shown in Fig. 13-17*b*. Observe that the resulting network is purely resistive and is therefore amenable to the analysis techniques developed in Part 2 of this book. Let the solution be given by

$$\begin{aligned} i_{L_j} &= I_j \qquad j = 1, 2, \ldots, k \\ v_{C_j} &= E_j \qquad j = (k + 1), (k + 2), \ldots, n \end{aligned} \tag{13-84}$$

If the pertinent state variables are i_{L_j} and v_{C_j}, then the equilibrium state is given by the point $(I_1, I_2, \ldots, I_k, E_{k+1}, E_{k+2}, \ldots, E_n)$. On the other hand, if the pertinent state variables are φ_j and q_j, then we can easily obtain

$$\varphi_j = \Phi_j(I_j) \qquad j = 1, 2, \ldots, k$$

from the φ_j-vs.-i_{L_j} curve and

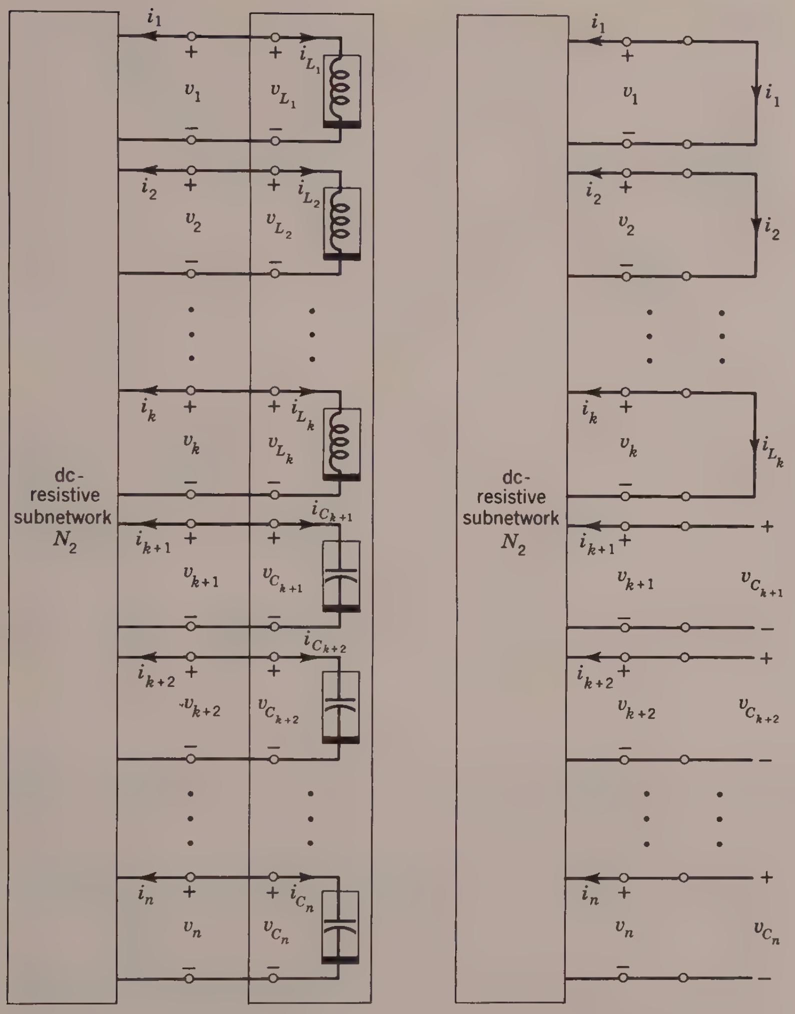
$$q_j = Q_j(E_j) \qquad j = (k + 1), (k + 2), \ldots, n$$

from the q_j-vs.-v_{C_j} curve. In this case the equilibrium state is given by the point $(\Phi_1(I_1), \Phi_2(I_2), \ldots, \Phi_k(I_k), Q_{k+1}(E_{k+1}), Q_{k+2}(E_{k+2}), \ldots, Q_n(E_n))$. The above interpretation is important enough to be stated as a theorem.

EXPLICIT EQUILIBRIUM STATE THEOREM

The equilibrium states of an autonomous network can be determined by solving for the short-circuit currents and open-circuit voltages of the *n*-port resistive network obtained by replacing each inductor by a short circuit and each capacitor by an open circuit.

This theorem reduces the problem of determining the equilibrium states of a dynamic network to the operating-point problem



(*a*) (*b*)

Fig. 13-17. The equilibrium state of an autonomous network can be obtained by solving for the terminal currents and voltages of the resistive network obtained by replacing each inductor by a short circuit and each capacitor by an open circuit.

of a resistive network. The principles developed in Part 2 of this book are therefore applicable here.

13-6-1 EQUILIBRIUM STATES OF A SWITCHING CIRCUIT: A NUMERICAL EXAMPLE

Consider the tunnel-diode switching circuit shown in Fig. 13-18*a*† and tunnel-diode *v-i* curve shown in Fig. 13-18*b*. The load-line construction shows three operating points, namely, $Q_1(v_1 = 0.06$ volt, $i = 0.76$ ma), $Q_2(v_1 = 0.29$ volt, $i = 0.6$ ma), and $Q_3(v_1 = 0.88$

---

† Except for the element values, this circuit was considered earlier in Fig. 6-5.

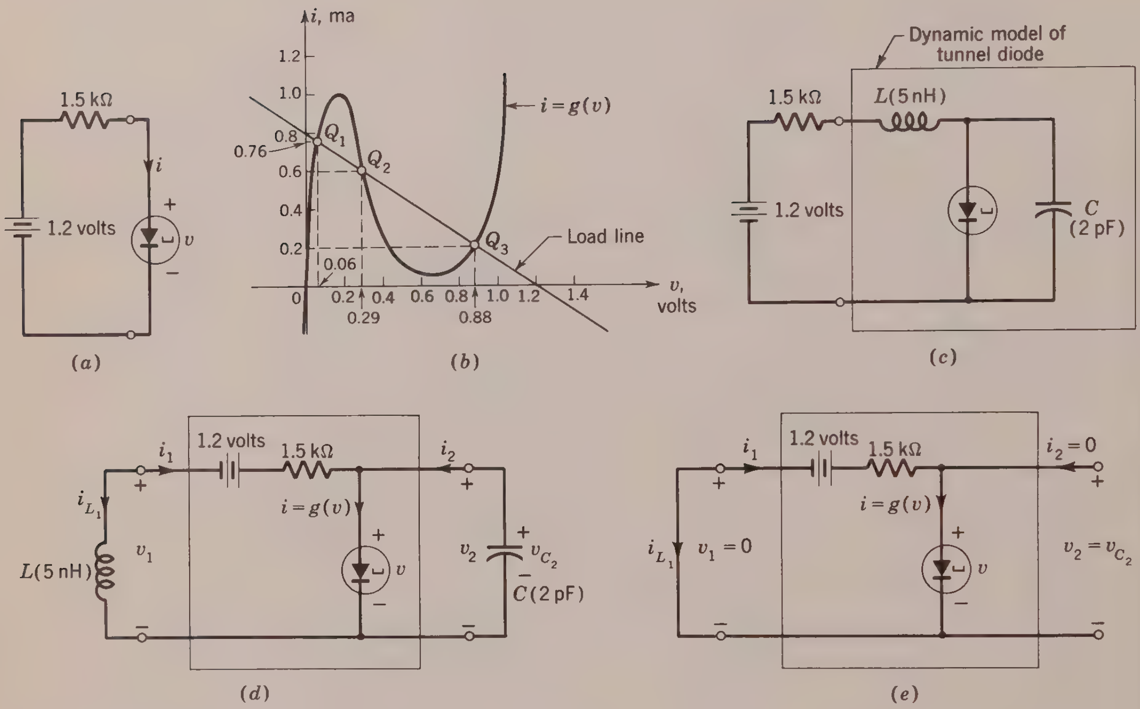


Fig. 13-18. To resolve the operating-point paradox, the dc-resistive tunnel-diode switching circuit is transformed into a second-order dynamic network by replacing the tunnel diode with a dynamic model.

volt, $i = 0.2$ ma). We know, of course, that only one operating point can be observed at any given time. The operating-point paradox (Sec. 5-5) questions the significance of these contradictory results. We shall now show that this paradox can be easily resolved by interpreting the operating point of a resistive network not as a physical solution but as an equilibrium state of an associated dynamic network obtained by including the parasitic elements. Once this step is taken, the operating-point paradox disappears because, as will be shown in the following section, not all equilibrium states are physically observable.

The dynamic network associated with this circuit can be easily obtained by connecting a parasitic inductance L and a parasitic capacitance C to the tunnel diode as shown in Fig. 13-18c. The three elements enclosed by the black box constitute a dynamic model of the tunnel diode.[1] This model is usually supplied by the manufacturer of the device with the pertinent parameter values. In this case, the parameters of the chosen tunnel diode are given by $L = 5$ nH (1 nanohenry $= 10^{-9}$ henry) and $C = 2$ pF (1 picofarad $= 10^{-12}$ farad). To find the equilibrium states of this second-order network, let us first redraw the network into the form of a two-port resistive network with the inductor and capacitor connected across ports 1 and 2, respectively, as shown in Fig. 13-18*d*. Let us find the equilibrium states of this network by two methods.

---

[1] The synthesis of an appropriate dynamic model for a physical device is a very difficult problem. Until a general theory is available, each device must be handled as a separate entity where the model is usually obtained from physical principles and intuitions.

Method 1 This method applies the implicit equilibrium state theorem and therefore requires that we derive the normal-form equation first. In order to do this analytically, we have approximated the *v-i* curve of the tunnel diode in Fig. 13-18*b* by the following polynomial[1]:

$$i = g(v) = 17.76v - 103.79v^2 + 229.62v^3 - 226.31v^4 + 83.72v^5 \quad \text{ma} \qquad (13\text{-}85)$$

The normal-form equations are easily obtained by the method described earlier in Sec. 13-4; namely,[2]

$$\frac{di_{L_1}}{dt} = \frac{-1.2 - (1.5)(10^3)i_{L_1} + v_{C_2}}{5(10^{-9})} \qquad (13\text{-}86)$$

$$\frac{dv_{C_2}}{dt} = \frac{-i_{L_1} - g(v_{C_2})(10^{-3})}{2(10^{-12})} \qquad (13\text{-}87)$$

where

$$g(v_{C_2}) = 17.76v_{C_2} - 103.79v_{C_2}{}^2 + 229.62v_{C_2}{}^3 - 226.31v_{C_2}{}^4 + 83.72v_{C_2}{}^5 \qquad (13\text{-}88)$$

We can now apply the implicit equilibrium state theorem and equate Eqs. (13-86) and (13-87) to zero.

$$-1.2 - 1.5(10^3)i_{L_1} + v_{C_2} = 0 \qquad (13\text{-}89)$$

$$-i_{L_1} - g(v_{C_2})(10^{-3}) = 0 \qquad (13\text{-}90)$$

To solve for i_{L_1} and v_{C_2}, we substitute Eq. (13-90) for i_{L_1} in Eq. (13-89).

$$-1.2 + 1.5(10^3)[g(v_{C_2})(10^{-3})] + v_{C_2} = 0 \qquad (13\text{-}91)$$

Substituting Eq. (13-88) for $g(v_{C_2})$ in Eq. (13-91) and simplifying, we obtain

$$125.58v_{C_2}{}^5 - 339.47v_{C_2}{}^4 + 344.43v_{C_2}{}^3 - 155.7v_{C_2}{}^2 + 27.7v_{C_2} - 1.2 = 0 \qquad (13\text{-}92)$$

---

[1] For details on how to approximate a curve by a polynomial, see Appendix A-3 and J. A. Narud and C. S. Meyer, A Polynomial Approximation for the Tunnel-Diode Characteristic, *IEEE Trans. Circuit Theory*, (Correspondence), vol. CT-10, p. 526, December, 1963.

[2] It is important that we maintain a consistent set of units throughout the derivation. For this example the units of i_{L_1} and v_{C_2} are amperes and volts. This means that the polynomial $g(v) = g(v_{C_2})$ in Eq. (13-85) must be multiplied by the factor 10^{-3}.

This is a fifth-degree polynomial equation and can be solved by numerical techniques. With the help of a computer, Eq. (13-92) is found to have three real solutions;[†] $v_{C_2} = 0.06, 0.29$, and 0.88 volts. Substituting these three values for v_{C_2} in Eq. (13-90), we obtain $i_{L_1} = -0.76, -0.6$, and -0.2 ma. The network, therefore, has the following three equilibrium states:

$$\begin{aligned} &\text{Equilibrium state 1:} \qquad && i_{L_1} = -0.76 \text{ ma} \qquad && v_{C_2} = 0.06 \text{ volt} \\ &\text{Equilibrium state 2:} && i_{L_1} = -0.6 \text{ ma} && v_{C_2} = 0.29 \text{ volt} \\ &\text{Equilibrium state 3:} && i_{L_1} = -0.2 \text{ ma} && v_{C_2} = 0.88 \text{ volt} \end{aligned} \tag{13-93}$$

Method 2 Let us now apply the explicit equilibrium state theorem by replacing the inductor by a short circuit and the capacitor by an open circuit, as shown in Fig. 13-18*e*. The equilibrium states, therefore, coincide with the (i_{L_1}, v_{C_2}) coordinates of the operating points of this resistive network. Observe that this circuit is identical with the original circuit in Fig. 13-18*a*. Hence, the same set of operating points obtained earlier by the load-line construction method constitutes the equilibrium states; namely,

$$\begin{aligned} &\text{Equilibrium state 1 } (Q_1)\text{:} \\ &\qquad i_{L_1} = -i_1 = -0.76 \text{ ma} \qquad v_{C_2} = v = 0.06 \text{ volt} \\ &\text{Equilibrium state 2 } (Q_2)\text{:} \\ &\qquad i_{L_1} = -i_1 = -0.6 \text{ ma} \qquad v_{C_2} = v = 0.29 \text{ volt} \\ &\text{Equilibrium state 3 } (Q_3)\text{:} \\ &\qquad i_{L_1} = -i_1 = -0.2 \text{ ma} \qquad v_{C_2} = v = 0.88 \text{ volt} \end{aligned} \tag{13-94}$$

A comparison between Eqs. (13-93) and (13-94) shows the same answers, as it should. Observe, however, that the second method is much easier to apply.

13-6-2 RELATIONSHIP BETWEEN TRAJECTORIES AND EQUILIBRIUM STATES

The solutions $x_1, x_2, \ldots, x_n$ of an nth-order autonomous network corresponding to some prescribed initial condition $(x_{1_0}, x_{2_0}, \ldots, x_{n_0})$ are generally functions of time $x_1 = X_1(t)$, $x_2 = X_2(t), \ldots, x_n = X_n(t)$. In most practical networks, these motions eventually settle down to some constant values $k_1, k_2, \ldots, k_n$. In this case the motion is said to be converging toward the equilibrium state $(k_1, k_2, \ldots, k_n)$. Since the motion of a network is completely de-

---

[†] For polynomial equations there exist several numerical techniques which will find all roots. Unfortunately, if the equation is not in polynomial form, then numerical techniques may fail to give all solutions.

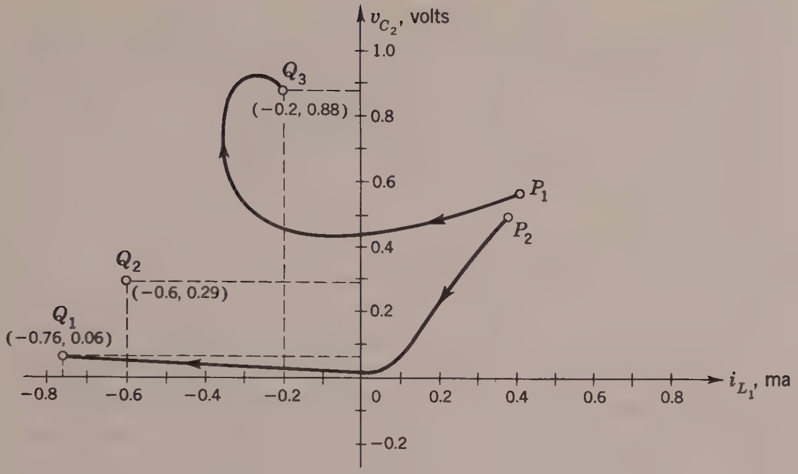


Fig. 13-19. Depending on the initial state, the trajectory may converge either toward Q_3 or Q_1.

scribed by its trajectory starting from the prescribed initial state, it follows that the associated trajectory must eventually arrive at the equilibrium state of the network.

If a network contains m equilibrium states, then depending on the initial state, the trajectory can converge toward only one of them. For example, let us consider the switching circuit shown in Fig. 13-18*c* again. The equilibrium states of this network are plotted in the i_{L_1}-vs.-v_{C_2} plane as shown in Fig. 13-19. By principles to be described in Chap. 19, it can be shown that the trajectory starting from the initial state P_1 converges toward the equilibrium state Q_3, whereas the trajectory starting from the initial state P_2 converges toward the equilibrium state Q_1. It will also be shown that no matter what the initial state is, no trajectory can ever converge toward the equilibrium state Q^2.† Observe that when the network is in equilibrium at Q_1 or at Q_3, the parasitic inductor can be replaced by a short circuit, and the parasitic capacitor by an open circuit. In either case the dynamic network of Fig. 13-18*d* reduces to the resistive network of Fig. 13-18*a*.

Using these observations, we can now conclude that although the original resistive network shown in Fig. 13-18*a* has three possible operating points, it will operate only at Q_1 or at Q_3, depending on the initial voltage across the parasitic capacitor.

Exercise 1: Consider the same switching circuit shown in Fig. 13-18*d*, but this time let both inductor and capacitor be nonlinear. If the inductor is characterized by $i_{L_1} = (10^{-7}\varphi_1)^{1/3}$ amp, and the capacitor is characterized by $v_{C_2} = (10^3 q_2)^{1/3}$ volts, find the equilibrium states in terms of the state variables φ_1 and q_2.

---

† The only exception here corresponds to the case where the initial state happens to be Q_2. In this case, the network remains at rest, and the corresponding trajectory degenerates to the point Q_2. This situation will never occur in practice because the slightest perturbation will cause the trajectory to leave Q_2.

Exercise 2: Find the equilibrium states of the network shown in Fig. 13-18*d* by determining the following two DP plots: (1) The DP plot across the inductor port with the capacitor replaced by an open circuit. (2) The DP plot across the capacitor port with the inductor replaced by a short circuit.

Exercise 3: Generalize the method used in Exercise 2 for finding the equilibrium states of an *n*th-order autonomous network.

13-7 STABILITY OF EQUILIBRIUM STATES

We have just seen an example of an equilibrium state which no trajectory can approach. Such an equilibrium state is often said to be unstable. A fundamental problem in the theory of dynamic networks is to determine the stability or instability of equilibrium states. This problem is of great practical significance because there are many physical systems which tend to become unstable unless properly designed. For example, consider a network containing only one equilibrium state. If this equilibrium state is unstable, then the trajectory of the network can be only a closed loop or an open trajectory which approaches a point at infinity.[1]

In the case of a closed-loop trajectory, otherwise known as a *limit cycle*,[2] the system will oscillate periodically. For example, a spacecraft with an unstable equilibrium state might tumble in space. This oscillatory motion is clearly undesirable in all cases except when the system is originally intended to be an oscillator. In the case of an open trajectory which tends towards infinity, the system will break down before this happens. For example, in an electronic circuit, some component will burn out as soon as its maximum voltage rating is exceeded.

In order to be able to avoid such catastrophic designs, it is necessary that we understand the concept of stability in sufficient depth. Intuitively, an equilibrium state of a physical system is said to be stable if, whenever the system is thrown off balance from an equilibrium position, it will eventually return to the original position; otherwise, the equilibrium state is said to be unstable. Before we formulate a precise definition of the stability of equilibrium states, it is instructive to consider the two simple systems shown in Fig. 13-20*a* and *b*. In both systems a metallic ball is constrained to move inside a glass tube with a prescribed contour.

It follows from the principles of mechanics that the equations of motion of these systems can be written in the normal form with the horizontal displacement x and the horizontal velocity component v_x as the state variables. Without writing down the normal-form equations, however, it is clear that any point where the ball

---

[1] The trajectory cannot converge toward a point in the finite plane because, by definition, such a point would qualify as another equilibrium state. This would contradict the assumption that the network has only one equilibrium state. Actually, the point at infinity can also be considered an equilibrium state. The behavior of the trajectories at infinity can be determined by performing an appropriate mathematical transformation. The details for carrying out this analysis can be found in more advanced books on *nonlinear oscillation*.

[2] Strictly speaking, all other trajectories near the limit cycle must, by definition, be open.

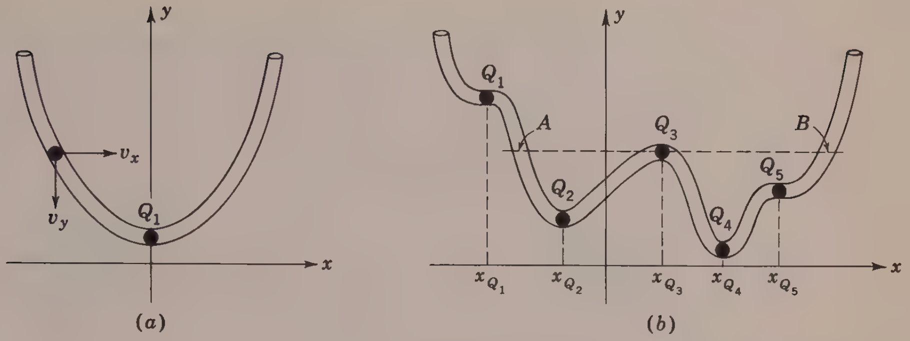


Fig. 13-20. Two physical systems for illustrating the concept of the stability of equilibrium states.

will remain at rest upon being released is an equilibrium state of the system. For example, for the system shown in Fig. 13-20*a*, there is only one equilibrium state, $Q_1(x = 0, v_x = 0)$. On the other hand, there are five equilibrium states in the system shown in Fig. 13-20*b*; namely, $Q_1(x = x_{Q_1}, v_x = 0)$, $Q_2(x = x_{Q_2}, v_x = 0)$, $Q_3(x = x_{Q_3}, v_x = 0)$, $Q_4(x = x_{Q_4}, v_x = 0)$, and $Q_5(x = x_{Q_5}, v_x = 0)$.

Although the above points all qualify as equilibrium states, they do not necessarily behave in the same manner when thrown off balance. In the case of Q_1 of Fig. 13-20*a* it is clear that no matter how far we perturb the ball from Q_1, the ball will always return to Q_1 eventually because of the friction between the ball and the glass tube. Hence, intuitively we say Q_1 is stable. A completely opposite behavior is observed at Q_1, Q_3, and Q_5 of Fig. 13-20*b*. Observe that the slightest perturbation of the ball from any of these points will result in the ball leaving that point forever. Hence, intuitively we say Q_1, Q_3, and Q_5 are unstable.

Consider next the remaining two equilibrium states. Observe that when perturbed, the ball at Q_2 will eventually return to Q_2, provided the initial perturbation is not large enough to force the ball above point *A*. Similarly, the ball at Q_4 will eventually return to Q_4, provided the initial perturbation does not carry the ball beyond point *B*. In either case we may say that the ball is stable, *provided* the initial perturbation is small. Observe that this condition is not needed in the case of Q_1 of Fig. 13-20*a*. In order to distinguish the two possible types of stable equilibrium states, we refer to Q_1 of Fig. 13-20*a* as being *stable-in-the-large* or *globally stable.* In contrast with this, we refer to Q_2 and Q_4 of Fig. 13-20*b* as *stable-in-the-small* or *locally stable.*

Let us consider the ideal case where the system shown in Fig. 13-20*a* is frictionless. Observe that if we repeat the above experiment, the ball will no longer return to its original position but will oscillate forever, passing through point Q_1 twice every cycle. According to our intuitive definition, we would say the equilibrium state Q_1 is unstable. However, this type of instability is very different from that of Q_1, Q_3, or Q_5 in Fig. 13-20*b*. Moreover, since the motion is periodic, there is a certain order or regularity that the other unstable equilibrium states did not possess. In view of this, many scientists prefer to call the above equilibrium state *stable,* while referring to any equilibrium state to which any trajectory eventually returns as being *asymptotically stable.* The above classification was first proposed by the Russian mathematician Liapunov and is generally accepted today as a reasonably useful definition for stability of equilibrium states. As such, we shall now summarize the above definitions in more precise terms.

13-7-1 STABILITY DEFINITION FOR SECOND-ORDER NETWORKS

Since the trajectories in a two-dimensional phase plane offer a much clearer geometrical interpretation, we shall formulate our stability definition in terms of second-order networks. The extension of this definition to *n*th-order networks will be given in the next section.

An equilibrium state Q with coordinates (x_{1_Q},x_{2_Q}) is said to be stable in the sense of Liapunov if all trajectories in the neighborhood of this point satisfy the following *thought experiment:* (1) Given *any* number $\epsilon > 0$, draw a circle C_ϵ with radius equal to ϵ and center at Q. (2) Find a smaller circle C_δ with radius $\delta > 0$ and also center at Q such that *any* trajectory which starts from an initial state *inside* this circle will never cross the bigger circle C_ϵ.†

The geometrical interpretation of this experiment is shown in Fig. 13-21*a*. The experiment fails in Fig. 13-21*b* because some trajectories starting inside C_δ would eventually cross C_ϵ. On the other hand, if all trajectories starting inside C_δ never cross the circle C_ϵ, as in Fig. 13-21*c*, then the experiment is satisfied for this *particular* value of ϵ. If it can be shown that the experiment is satisfied for *all* values of ϵ, then the system is stable. Notice that in order to be stable, the trajectories need not converge toward the equilibrium state Q. They could just as well form closed curves which correspond to periodic solutions.

If in addition to being stable all trajectories eventually approach the equilibrium state Q, then the network is said to be *asymptotically stable.* A typical trajectory satisfying this condition is shown in Fig. 13-21*d*.

Notice that although the above stability definitions agree with

---

† This thought experiment must be carried out for all values of ϵ, and corresponding to each circle C_ϵ, a *nonzero* radius δ must be found which has the desired properties.

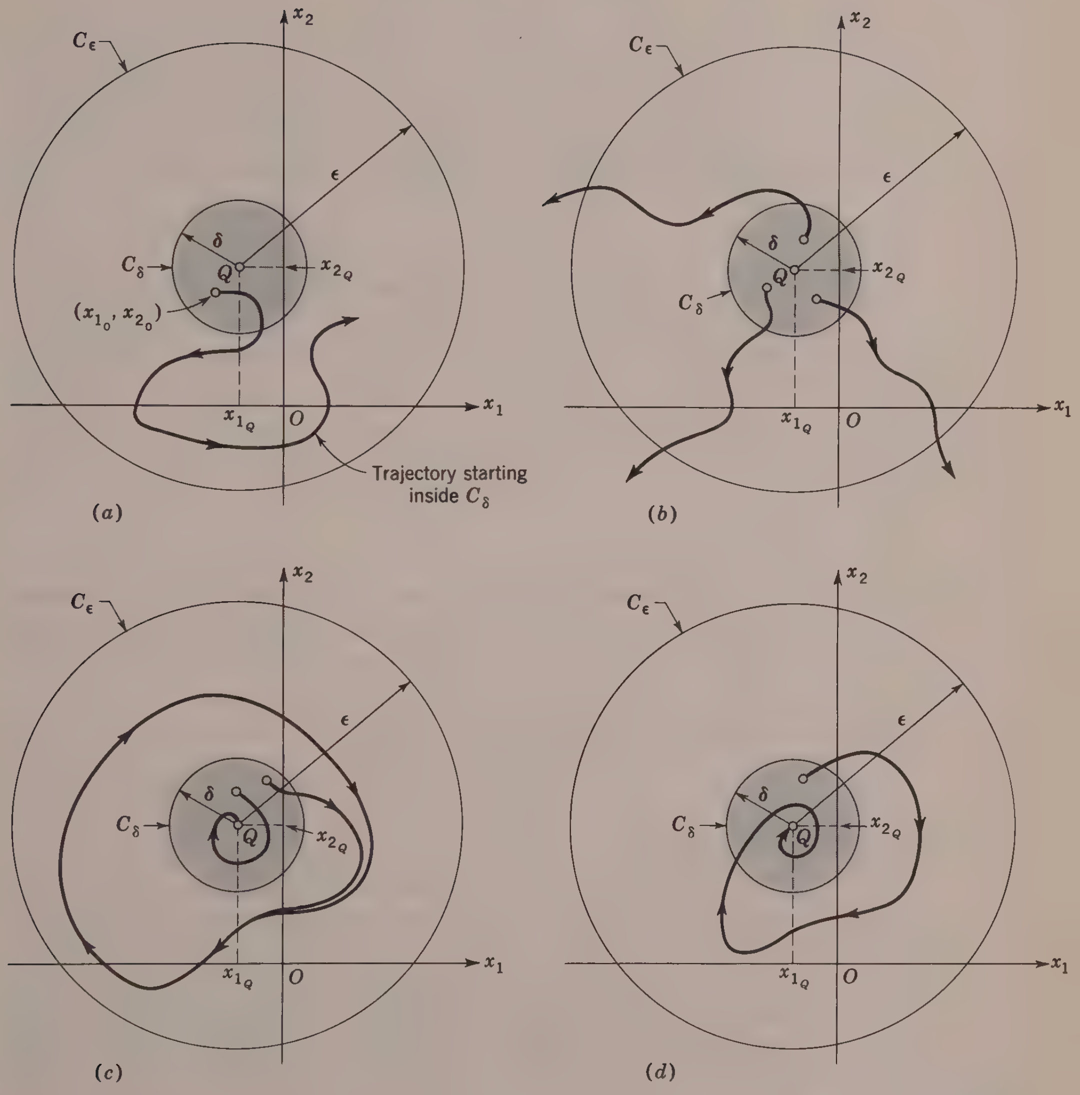


Fig. 13-21. A geometrical interpretation of various stability definitions.

our intuition, they are very difficult to verify. This is due to the requirement that the experiment be satisfied for all possible values of ϵ. One certainly cannot afford to test each and every value of ϵ. On the other hand, it is usually much easier to show that a network is unstable. This can be done by exhibiting only one trajectory which fails the test. In view of the difficulty for showing stability, we shall formulate some useful stability criteria in the following chapters.

13-7-2 STABILITY DEFINITION FOR *n*TH-ORDER NETWORKS

We can now generalize the above stability definitions to *n*th-order networks. Instead of circles with radius ϵ and δ, we now talk of *n*-dimensional spheres with radius ϵ and δ. An *n*-dimensional sphere S_ϵ with radius ϵ is defined to be the set of all points whose distance from the equilibrium state is equal to ϵ. Our stability definition now takes the following form: An equilibrium state Q in an *n*-dimensional state space is *stable in the sense of Liapunov* if, given any value $\epsilon > 0$ and any initial time $t = t_0$, it is possible to find a corresponding sphere S_δ with $0 < \delta < \epsilon$ such that all trajectories starting inside S_δ will never cross the larger sphere S_ϵ. If, in addition to this property, all trajectories starting inside S_δ eventually approach Q, then the equilibrium state is asymptotically stable.

Exercise 1: Define stability and asymptotic stability for a first-order network.

Exercise 2: Define stability and asymptotic stability for a second-order network.

Exercise 3: Explain why all stability definitions require that the prescribed tests be satisfied for all values of ϵ, and not just for one.

Exercise 4: Sketch a geometrical interpretation of the stability and asymptotic stability definitions for third-order systems.

13-8 SUMMARY

Order of complexity This is defined as the maximum number of independent initial conditions that can be specified in terms of the electrical variables in the network. It is also equal to the number of state variables in the normal-form equations associated with the network. This number corresponds, therefore, to the minimum number of initial conditions that must be specified in order to solve the network completely, that is, without the need for introducing arbitrary constants.

Classification of dynamic nonlinear networks Any network can be classified as a zero-order network, a first-order network, a second-order network, . . . , or as an *n*th-order network. A zero-order network is simply a synonym for resistive networks. All other classes are dynamic networks.

Significance of initial conditions Initial conditions are introduced in lieu of the past history of the voltages and currents in the energy-storage elements, which are usually unknown. Some appropriate sets of initial conditions are the "charge" or "voltage" of a capacitor and the "current" or "flux linkage" of an inductor. The order of complexity of *RLC* networks can be determined by inspection. See Eq. 13-10.

Alternate method for deriving normal-form equations This method extracts all energy-storage elements from the network and derives the equations in terms of the port description of the resistive subnetwork.

Trajectories These consist of space curves (not a surface) representing the solution waveforms in the state space with the variable time as a parameter.

Equilibrium states The points on the state space for which the solution $dx_j/dt \equiv 0$, $j = 1, 2, \ldots, n$. There are two methods for determining the equilibrium states:

1. Implicit equilibrium state theorem. Obtained by equating the right-hand side of the normal-form equations to zero and solving these equations by numerical methods.
2. Explicit equilibrium state theorem. Obtained by finding the operating points when all inductors are replaced by short circuits and all capacitors are replaced by open circuits.

Stability A network may be stable, asymptotically stable, or unstable. Roughly speaking, it is stable if the trajectories do not tend to infinity. It is asymptotically stable if all trajectories converge to the equilibrium state. It is unstable if it is not stable.

PROBLEMS

13-1 Find the order of complexity of the networks shown in Fig. P13-1*a* to *d* by inspection.

13-2 A certain mechanical spring-mass system under stress can be modeled by the dynamical network shown in Fig. P13-2. The in-

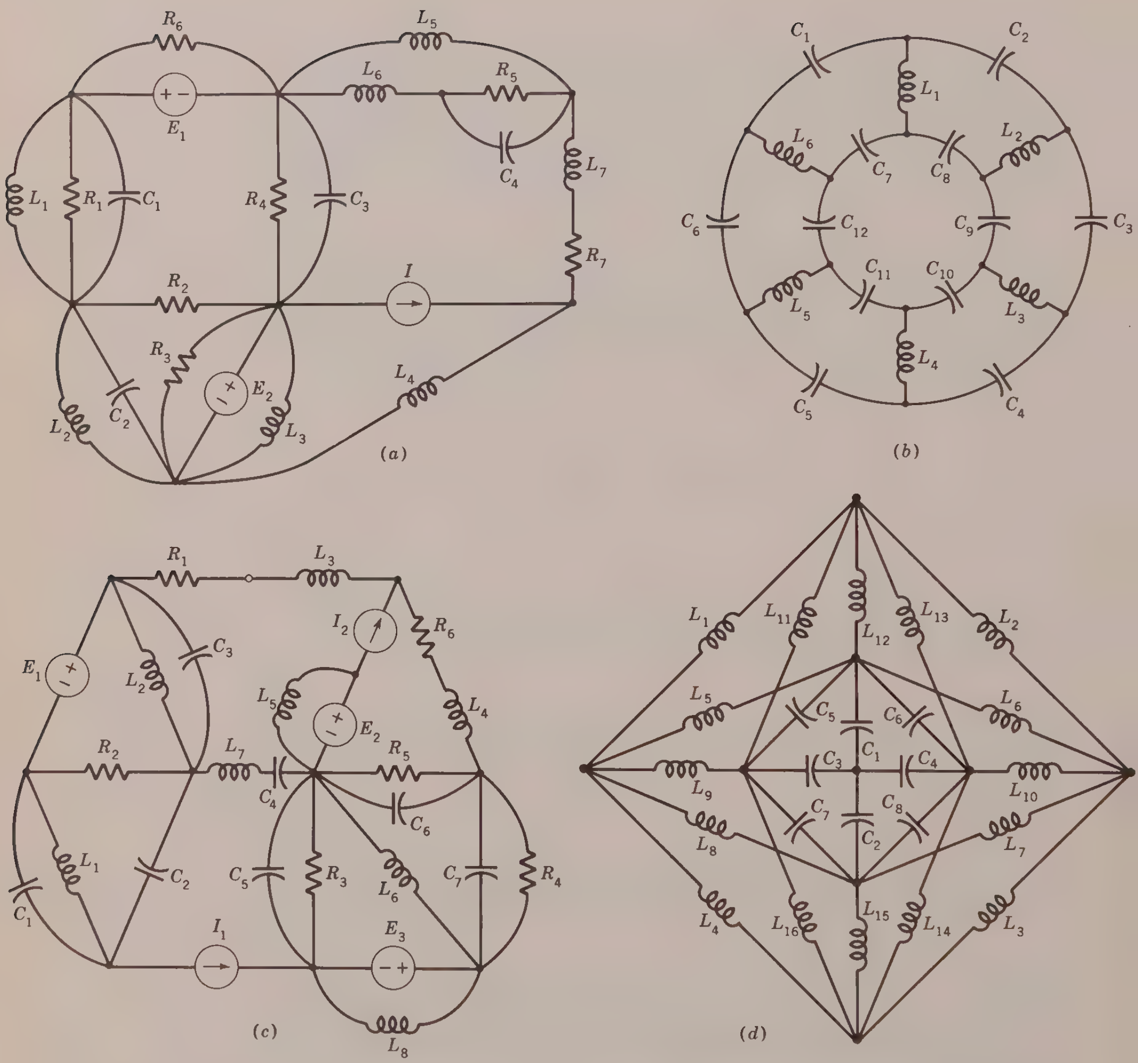


Fig. P13-1.

dependent sources are time-dependent in order to simulate the externally applied forces.

(a) Find the order of complexity of this mechanical system.

(b) With capacitor charges q_j and inductor flux linkages φ_j chosen to be the state variables, write down the general functional form of the equations of motion in the normal form.

(c) Is this an autonomous or a nonautonomous network? Is the concept of an equilibrium state applicable here? Explain why.

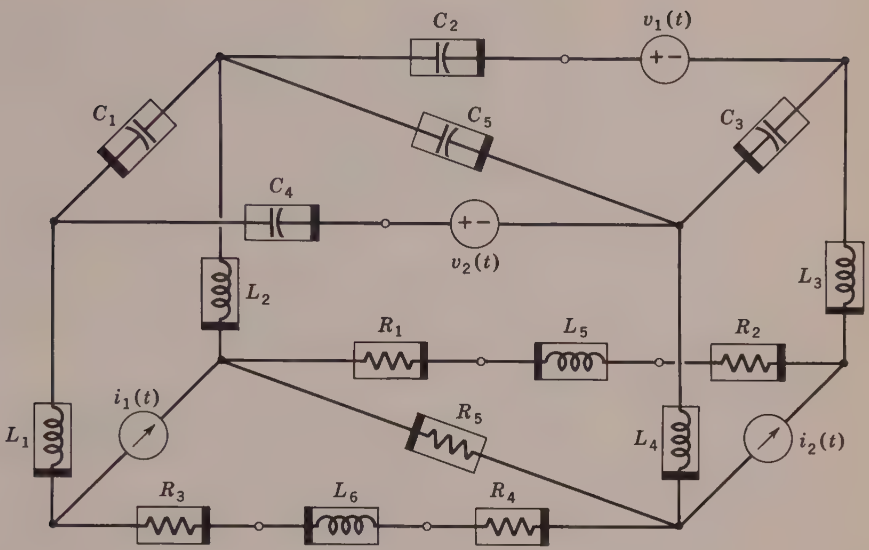
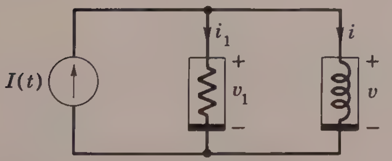


Fig. P13-2.

13-3 Consider the network shown in Fig. P13-3. With i as the state variable, derive the equations of motion in the normal form by two methods:

(a) By the elimination and substitution method developed in Sec. 4-6.

(b) By finding the DP plot across the resistive one-port network described in Sec. 13-4.



Element characteristics:

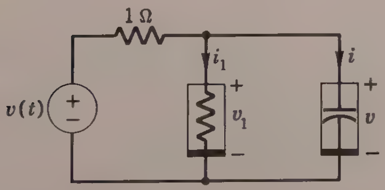
Resistor: $v_1 = i_1 + i_1^3$

Inductor: $\varphi = \tanh i$

Current source: $I(t) = 10 \cos \omega t$

Fig. P13-3.

13-4 Repeat Prob. 13-3 for the network shown in Fig. P13-4 with v as the state variable.



Element characteristics:

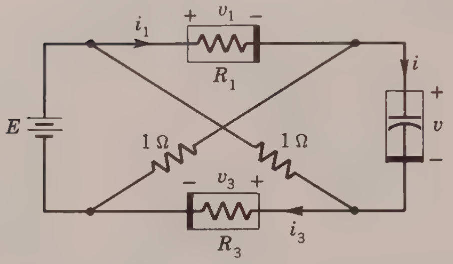
Resistor: $i_1 = v_1^{1/2}, \quad v_1 \geq 0$

$= -|v_1|^{1/2}, \quad v_1 < 0$

Capacitor: $q = v^3 + v + 2$

Fig. P13-4.

13-5 Derive the normal-form equation of the network shown in Fig. P13-5 in terms of the state variable q. HINT: Apply the v-shift theorem and make use of the standard formula for solving cubic equations given in most mathematical tables and handbooks.



Element characteristics:

Resistor R_1: $i_1 = v_1^3$

Resistor R_3: $i_3 = v_3^3$

Capacitor: $v = 2q^3 + q^2 - 1$

Fig. P13-5.

13-6 Consider the network shown in Fig. P13-6.

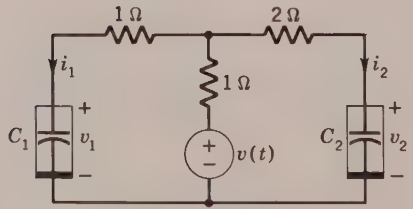
(a) Find the normal-form equations in terms of the state variables q_1 and q_2.

(b) With $v_1 = 4q_1$, $v_2 = 8q_2$, and $v(t) = 4$ volts, determine the equilibrium states by applying the implicit equilibrium state theorem.

(c) Repeat (*b*) by applying the explicit equilibrium state theorem.

(d) Repeat (*a*) in terms of the state variables v_1 and v_2.

(e) Repeat (*a*) in terms of the state variables q_1 and v_2.



Element characteristics:

C_1: $v_1 = f_1(q_1)$ or $q_1 = f_1^{-1}(v_1)$

C_2: $v_2 = f_2(q_2)$ or $q_2 = f_2^{-1}(v_2)$

Fig. P13-6.

13-7 Consider the network shown in Fig. P13-7.

(a) Find the normal-form equations in terms of the state variables φ_1 and φ_2.

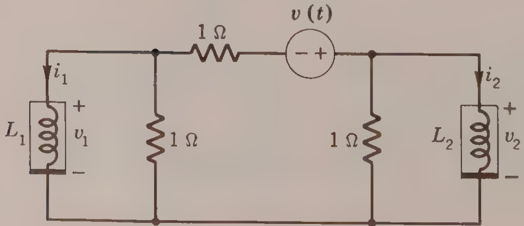
(b) With $\varphi_1 = \frac{1}{3}i_1$, $\varphi_2 = \frac{1}{3}i_2$, and $v(t) = 5$ volts, determine the equilibrium states by applying the implicit equilibrium state theorem.

(c) Repeat (*b*) by applying the explicit equilibrium state theorem.

(d) Repeat (*a*) in terms of the state variables i_1 and i_2.

(e) Repeat (*a*) in terms of the state variables φ_1 and i_2.

Fig. P13-7.



Element characteristics:

L_1: $\varphi_1 = f_1(i_1)$ or $i_1 = f_1^{-1}(\varphi_1)$

L_2: $\varphi_2 = f_2(i_2)$ or $i_2 = f_2^{-1}(\varphi_2)$

13-8 Consider the network shown in Fig. P13-8.

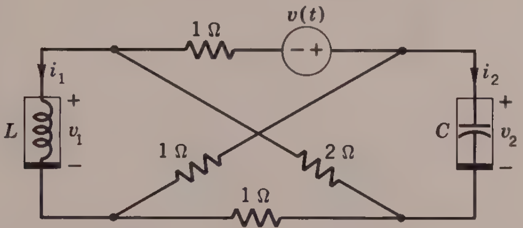
(a) Find the normal-form equations in terms of the state variables φ_1 and q_2.

(b) With $\varphi_1 = \frac{1}{3}i_1$, $q_2 = \frac{1}{2}v_2$, and $v(t) = 5$ volts, find the equilibrium states by applying the implicit equilibrium state theorem.

(c) Repeat (*b*) by applying the explicit equilibrium state theorem.

(d) Repeat (*a*) in terms of the state variables i_1 and v_2.

(e) Repeat (*a*) in terms of the state variables φ_1 and v_2.



Element characteristics:

L: $i_1 = f_1(\varphi_1)$ or
$\varphi_1 = f_1^{-1}(i_1)$

C: $q_2 = f_2(v_2)$ or
$v_2 = f_2^{-1}(q_2)$

Fig. P13-8.

13-9 Find the normal-form equations for the networks shown in Fig. P13-9*a* to *c*. Compare the order of complexity of each network with that predicted by the theorem on the order of complexity of dynamic *RLC* networks in Sec. 13-2-2. Explain any discrepancy that may arise.

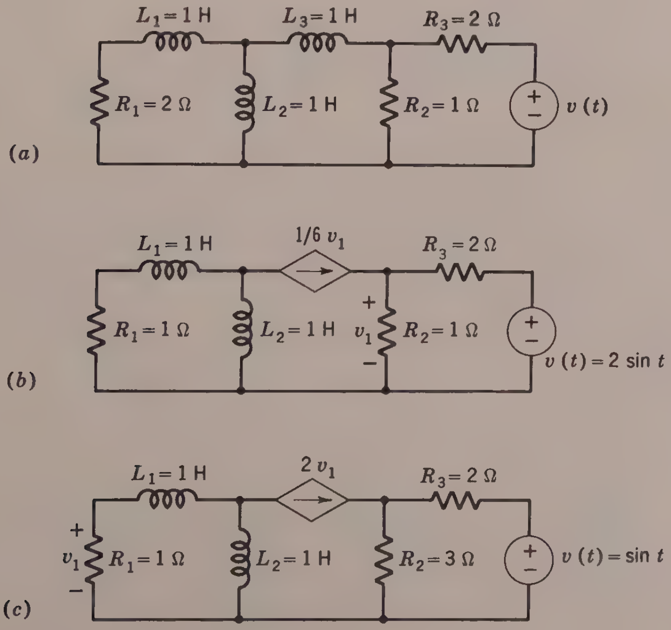


Fig. P13-9.

13-10 Find the normal-form equations for the networks shown in Fig. P13-10*a* to *c*. Compare the order of complexity of each network with that predicted by the theorem on the order of complexity of dynamic *RLC* networks in Sec 13-2-2. Explain any discrepancy that may arise.

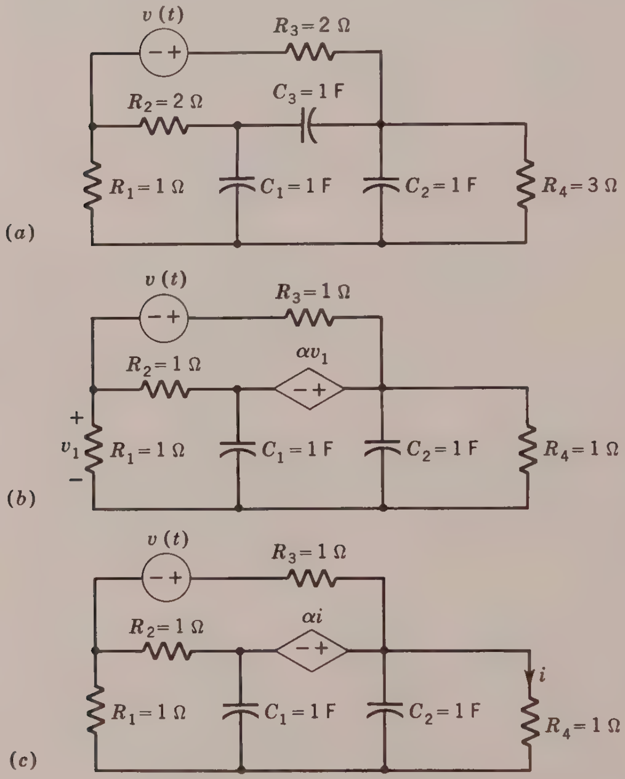
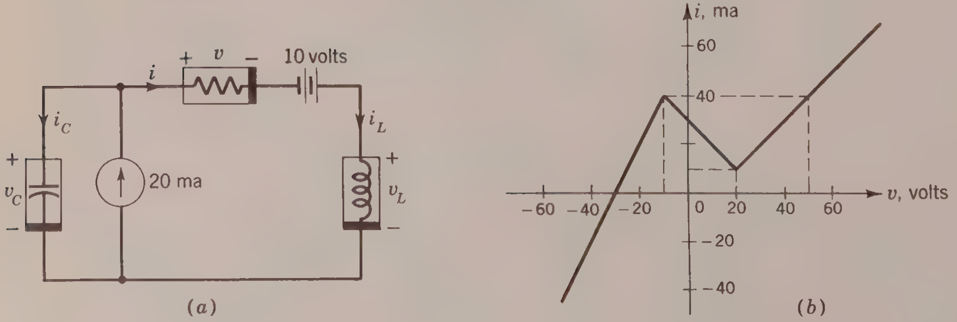


Fig. P13-10.

13-11 Consider the nonlinear network shown in Fig. P13-11*a*, where the nonlinear resistor, inductor, and capacitor are characterized by the *v*-*i*, *i*-φ, and *v*-*q* curves shown in Fig. P13-11*b* to *d*, respectively. Find the equilibrium states by applying the explicit equilibrium state theorem.

Fig. P13-11.



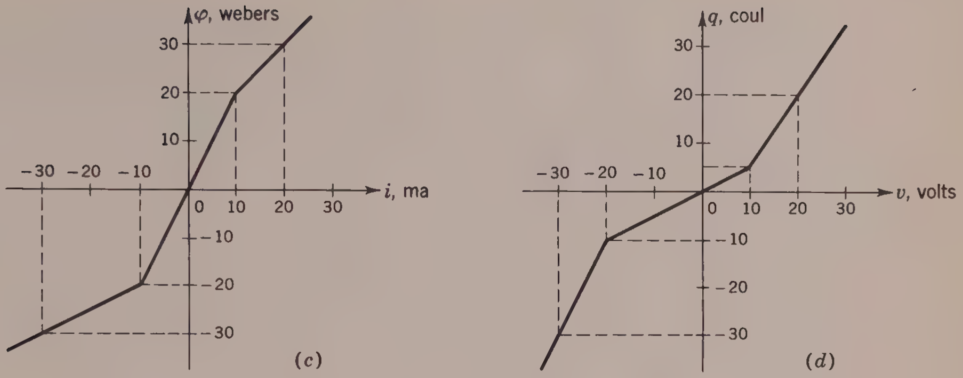


Fig. P13-11. (*Continued*)

13-12 Consider the network shown in Fig. P13-12.

(a) Find the normal-form equation in terms of the state variable i_1.

(b) Let $v(t) = e^{-t}$ and solve for $i_1(t)$ with the initial condition $i_1(0) = 5$. Sketch the trajectory of solution.

(c) Repeat (*b*) with $v(t) = \sin t$.

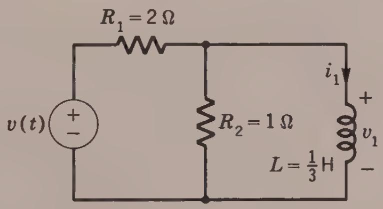


Fig. P13-12.

13-13 The normal-form equation of a second-order linear network assumes the general form

$$\frac{dx_1}{dt} = a_{11}x_1 + a_{12}x_2 + k_1$$

$$\frac{dx_2}{dt} = a_{21}x_1 + a_{22}x_2 + k_2$$

(a) Find the equilibrium states in terms of the parameters a_{ij} and k_j.

(b) Find the solutions $x_1(t)$ and $x_2(t)$ corresponding to each of the following sets of parameters:

$a_{11} = a_{22} = -1,\ a_{12} = a_{21} = 0,\ k_1 = k_2 = 0$

$a_{11} = a_{22} = 1,\ a_{12} = a_{21} = 0,\ k_1 = k_2 = 0$

$a_{11} = -2,\ a_{22} = -1,\ a_{12} = a_{21} = 0,\ k_1 = k_2 = 0$

$a_{11} = a_{22} = -1,\ a_{12} = 0,\ a_{21} = 1,\ k_1 = k_2 = 0$

(c) Draw the trajectories corresponding to the solutions in (*b*) with the initial conditions $(x_1(0), x_2(0)) = (5,5), (0,5), (-5,5), (-5,0), (-5,-5), (0,-5), (-5,-5)$, and $(5,0)$, respectively.

(d) Based on the trajectories in (*c*), classify the equilibrium state (0,0) according to the type of stability it possesses, namely, stable, asymptotically stable, or unstable. State whether the classification is valid only locally or globally.

13-14 Repeat parts (*b*), (*c*), and (*d*) of Prob. 13-13 for the following sets of parameters:

$$a_{11} = -1,\ a_{22} = 1,\ a_{12} = a_{21} = 0,\ k_1 = k_2 = 0$$

$$a_{11} = a_{22} = -1,\ a_{12} = -1,\ a_{21} = 1,\ k_1 = k_2 = 0$$

$$a_{11} = a_{22} = 1,\ a_{12} = -1,\ a_{21} = 1,\ k_1 = k_2 = 0$$

$$a_{11} = a_{22} = 1,\ a_{12} = 1,\ a_{21} = -1,\ k_1 = k_2 = 0$$

$$a_{11} = a_{22} = 0,\ a_{12} = -1,\ a_{21} = 1,\ k_1 = k_2 = 0$$

13-15 Let x be the state variable associated with a first-order network.

(a) If the solution $x(t)$ is as shown in Fig. P13-15*a*, plot the associated trajectory.

(b) If the trajectory is given by Fig. P13-15*b*, plot the corresponding solution $x(t)$.

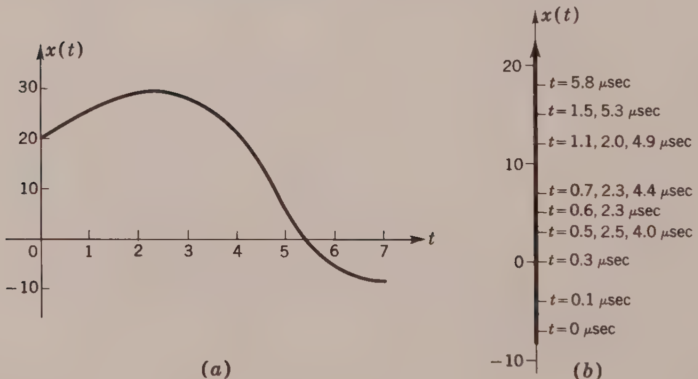
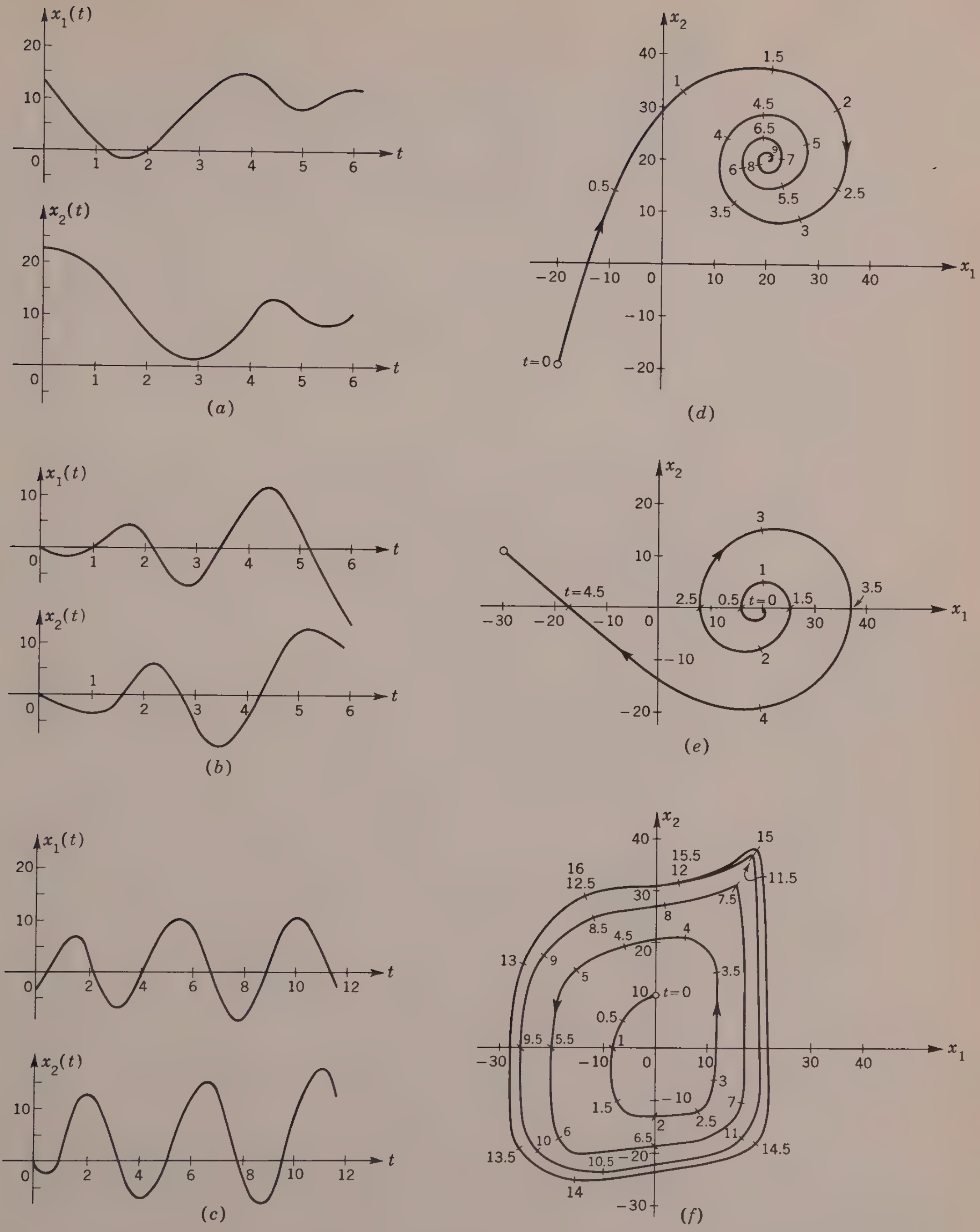


Fig. P13-15. (*a*) (*b*)

13-16 Let x_1 and x_2 be the state variables associated with a second-order network.

(a) Plot the trajectories associated with the three sets of solutions $(x_1(t), x_2(t))$ as shown in Fig. P13-16*a* to *c*.

(b) Plot the solutions $(x_1(t), x_2(t))$ corresponding to each of the trajectories shown in Fig. P13-16*d* to *f*.



(a)
$x_1(t)$
$x_2(t)$
t
0
1
2
3
4
5
6
10
20
(b)
(c)
8
10
12
(d)
x_2
x_1
t=0
0.5
1
1.5
2
2.5
3
3.5
4
4.5
5
5.5
6
6.5
7
8
9
−20
−10
30
40
(e)
t=4.5
t=0
0.5
1.5
2.5
3.5
−30
(f)
7.5
8.5
9.5
10.5
11
11.5
12
12.5
13
13.5
14
14.5
15
15.5
16

Fig. P13-16.

14 ANALYSIS OF AUTONOMOUS FIRST-ORDER NONLINEAR NETWORKS

14-1 BASIC PHILOSOPHY AND APPROACH

The normal-form equation of a first-order autonomous network has been shown to be of the form

$$\frac{dx}{dt} = f(x) \tag{14-1}$$

The expression $f(x)$ can be obtained from the DP plot of the associated one-port resistive subnetwork and may either be a single-valued or a multivalued function of the state variable x.† If $f(x)$ is easily represented in analytic form, then the numerical techniques, such as those presented in Sec. 4-7-1, can be used to find the solution of Eq. (14-1) with the help of a computer. One disadvantage of this approach, however, is that it does not give much insight into the network's behavior. This is because anyone who has a computer program for solving Eq. (14-1) can find the solution corresponding to any given initial condition without the slightest idea of what a capacitor or an inductor inside the network does. Actually, in order to make good use of a computer, one must have some idea of the general behavior of the network first, rather than resort to a blind use of the computer. In fact, whenever one is confronted with the analysis or synthesis of a new network, a qualitative analysis is much more meaningful than a quantitative analysis. Some of the important qualitative results that are useful to know are the location and stability of the equilibrium states. Another desirable piece of qualitative information is the behavior of the trajectories near an equilibrium state.

In view of the above considerations, we shall present first some useful techniques for obtaining a qualitative analysis of any

† Recall that $f(x)$ is a single-valued function of x if for each value $x = x_k$, there corresponds one and only one value $f(x_k)$. In this case we say $f(x)$ is *x-controlled.*

first-order autonomous network. Once the qualitative analysis is complete, the next logical step will be to perform a quantitative analysis with some meaningful (rather than arbitrary) choice of initial conditions. The piecewise-linear approach will be introduced for obtaining the desired quantitative results.

14-2 EQUILIBRIUM STATES AND STABILITY CRITERIA

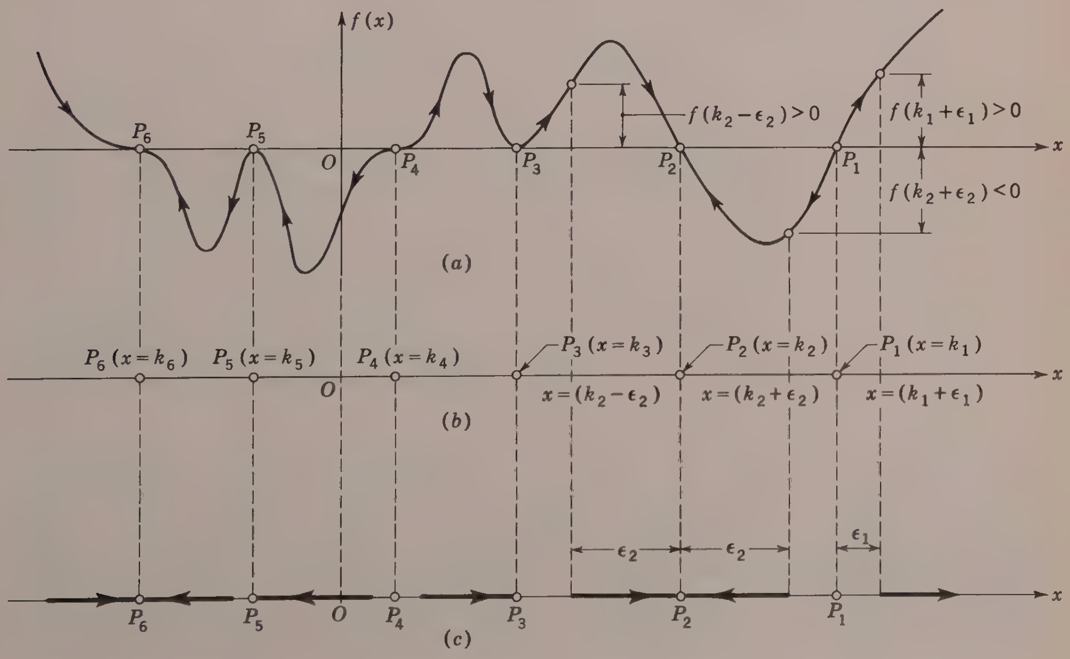
The equilibrium states corresponding to Eq. (14-1) are given by the solutions of

$$f(x) = 0 \tag{14-2}$$

Graphically, these consist of the points of intersection of the curve $f(x)$ with the x axis. In order to investigate the stability of these equilibrium states, let us consider the hypothetical curve $f(x)$ as shown in Fig. 14-1*a*. The equilibrium states of this network are obtained by projecting the x coordinate of the points of intersections of $f(x)$ with the x axis as shown in Fig. 14-1*b*.

To determine the stability of each equilibrium state P_j, it is necessary to perform a conceptual experiment around the neighborhood of P_j in accordance with Liapunov's definition. This con-

Fig. 14-1. The equilibrium states of a first-order network are given by the x coordinates of the points of intersections of $f(x)$ with the x axis.



sists of perturbing the state variable x a distance ϵ to the right or to the left of P_j,† and observing the resulting trajectory starting from this new initial state. If the trajectory corresponding to *each* value of ϵ always returns to P_j, the equilibrium state P_j is stable. On the other hand, if there is a value of ϵ for which the trajectory diverges from P_j, the equilibrium state P_j is unstable.

Consider first the equilibrium state P_1. Suppose we begin by perturbing the state variable x from $x = k_1$ to $x = (k_1 + \epsilon_1)$ at an arbitrary initial time $t = t_0$. Our problem is to investigate the nature of the solution $x = X(t)$ with $X(t_0) = (k_1 + \epsilon_1)$ as the initial condition. At this point $f(x) = f(k_1 + \epsilon_1) > 0$ (Fig. 14-1*a*). Since the solution $x = X(t)$ must satisfy Eq. (14-1) for all time t, we have at $t = t_0$

$$\frac{dx}{dt} > 0$$

This implies that the solution $x = X(t)$ starting from $x(t_0) = (k_1 + x_1)$ must be an increasing function of time at $t = t_0$, and therefore the corresponding trajectory must be moving to the right of P_1, as shown in Fig. 14-1*c*. Moreover, since $f(x) > 0$ for $x > (k_1 + \epsilon_1)$, it is clear that the trajectory must keep on moving to the right and will never return to P_1. Hence, we conclude that the equilibrium state P_1 is unstable.

Consider next the equilibrium state P_2. If we displace x from $x = k_2$ to $x = (k_2 + \epsilon_2)$, as shown in Fig. 14-1*c*, we find that $dx/dt < 0$, and the trajectory must be moving toward P_2. This behavior is easily seen to be true for *any* other value of ϵ_2 that results in an initial state between P_2 and P_1. However, we cannot conclude that P_2 is stable because we have not yet tested the case where x is displaced to the left of P_2. Hence let us perturb x from $x = k_2$ to $x = (k_2 - \epsilon_2)$. This time we have $f(k_2 - \epsilon_2) > 0$, and the trajectory is again seen to move toward P_2. Since this behavior is true for *any* value of ϵ_2 that results in an initial state between P_2 and P_3, we conclude that the equilibrium state P_2 is stable.

Consider next the equilibrium state P_3. It is seen from Fig. 14-1*a* that any trajectory corresponding to any initial state between P_3 and P_4 will converge toward P_3. However, the trajectory corresponding to any initial state between P_3 and P_2 tends to diverge from P_3. Therefore, we conclude that P_3 is unstable.

By a similar procedure we can readily establish that the equilibrium states P_4 and P_5 are also unstable, and that the equilibrium state P_6 is stable. Hence, of the six equilibrium states, only P_2 and P_6 are stable. If we examine Fig. 14-1*a* for $f(x)$ at these two points more carefully, we observe that $f(x)$ is a monotonically decreasing function within a small interval centered about P_2 and about P_6.† In contrast to this we see that within the interval about each unstable equilibrium state, $f(x)$ is not a monotonically decreasing function of x. These observations allow us to state the following stability criteria.

---

† By definition, the perturbation ϵ can be any arbitrary number, provided it does not cause the resulting initial state to fall outside of the equilibrium states adjacent to P_j.

STABILITY CRITERIA FOR FIRST-ORDER NETWORKS

Let $dx/dt = f(x)$ be the normal-form equation of a first-order network, where $f(x)$ is a continuous single-valued function of x. Let $P_j(x = k_j)$ be an equilibrium state of the network. Then P_j is stable if there is an interval around $x = k_j$ where $f(x)$ is a monotonically decreasing function of x. Otherwise, P_j is unstable.

Applying this theorem to the function $f(x)$ shown in Fig. 14-1*a*, we can conclude immediately by inspection that P_2 and P_6 are stable, whereas P_1, P_3, P_4, and P_5 are unstable. The equilibrium states P_1 and P_4 are unstable because there is no interval about P_1 or P_4 where $f(x)$ is monotonically decreasing. The equilibrium state P_3 is unstable even though $f(x)$ is monotonically decreasing on the left of P_3 because $f(x)$ becomes monotonically increasing on the right of P_3. Similarly, although $f(x)$ is monotonically decreasing on the right of P_5, it is monotonically increasing on the left of P_5; therefore, P_5 is unstable.

If we take another look at Fig. 14-1*a*, we observe that the slope of $f(x)$ at each equilibrium state P_j may either be zero, positive, or negative. The slope will be zero if, and only if, P_j is a minimum, a maximum, or a point of inflection of $f(x)$. In most practical problems the slope of $f(x)$ at P_j is either positive or negative but not zero. If the slope is positive, $f(x)$ is monotonically increasing, and if the slope is negative, $f(x)$ is monotonically decreasing. Therefore, we can state a corollary to the theorem of stability criteria for first-order networks.

COROLLARY TO STABILITY CRITERIA FOR FIRST-ORDER NETWORKS

If the slope of $f(x)$ is nonzero at each equilibrium state P_j, then P_j is stable if the slope of $f(x)$ at P_j is negative and P_j is unstable if the slope of $f(x)$ at P_j is positive. Moreover, if $f(x)$ is a single-valued function, the equilibrium states must alternate between stability and instability.

---

† A function $f(x)$ is said to be monotonically decreasing if $f(x_2) \leq f(x_1)$ whenever $x_2 > x_1$.

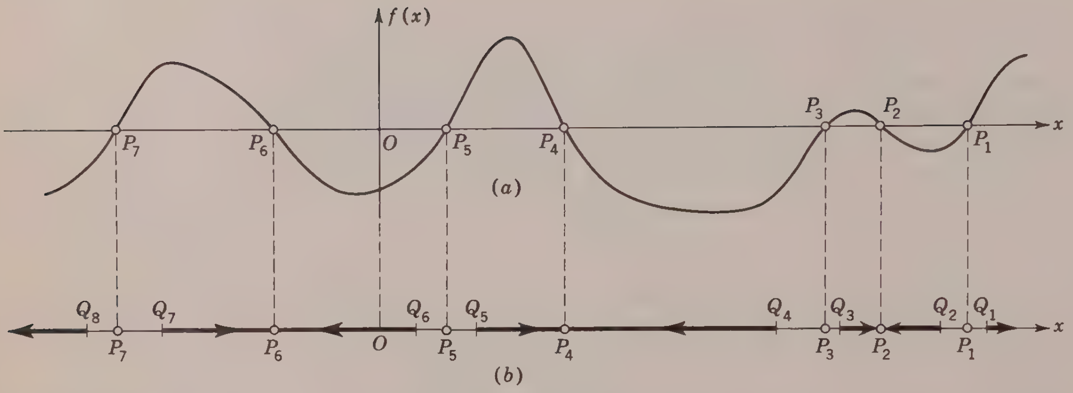


Fig. 14-2. An example illustrating the determination of the stability of equilibrium states by inspection. An equilibrium state P_j is stable if the slope of $f(x)$ at P_j is negative; otherwise, it is unstable.

To demonstrate the utility of this corollary, consider the curve $f(x)$ as shown in Fig. 14-2*a*. Since the slopes at P_1, P_3, P_5, and P_7 are positive, these equilibrium states are unstable. On the other hand, the equilibrium states P_2, P_4, and P_6 are stable because the slopes of $f(x)$ at these points are negative. Alternately, we could have obtained the same conclusion by determining the stability of any one equilibrium state and then using the alternating property to determine the stability of the remaining equilibrium states. For example, let us arbitrarily choose P_3 and examine its stability. Since the slope of $f(x)$ at P_3 is positive, we conclude that P_3 is unstable. From this information alone we can immediately conclude that the equilibrium states P_1, P_5, and P_7 must be unstable, whereas the equilibrium states P_2, P_4, and P_6 must be stable.

Once we find the stability of each equilibrium state of a first-order network, we can draw the trajectory corresponding to any initial state. For example, the trajectories corresponding to the initial states $Q_1, Q_2, \ldots, Q_8$ are easily obtained as shown in Fig. 14-2*b*. The trajectory corresponding to the initial state Q_1 must diverge from P_1 because the equilibrium state P_1 is unstable. On the other hand, the trajectories corresponding to the initial states Q_4 and Q_5 must both converge toward P_4 because the equilibrium state P_4 is stable. If we can also find the time parameter associated with each trajectory, as will be considered in the following section, then we can recover the solution $x = X(t)$ corresponding to any initial condition $X(t_0) = x_0$.†

14-3 STABILITY CRITERIA FOR TWO COMMON CIRCUIT CONFIGURATIONS

Let us now apply the stability criteria of the preceding section to derive a simple criterion for two commonly encountered circuit configurations.

---

† It must be emphasized, however, that in many practical situations we do not even have to do this because we are usually interested only in the general properties of the solutions and not in the exact solution waveforms.

Configuration 1 Consider the network shown in Fig. 14-3*a*. If we denote the v_R-i_R curve of the nonlinear resistor by $i_R = g(v_R)$, then the normal-form equation is easily found to be

$$\frac{dv_C}{dt} = \frac{1}{C}\left[\frac{E - v_C}{R} - g(v_C)\right] \tag{14-3}$$

Comparing this equation with Eq. (14-1), we obtain

$$f(v_C) = \frac{1}{C}\left[\frac{E - v_C}{R} - g(v_C)\right] \tag{14-4}$$

where the state variable x is now given by $v_C = v_R$. The equilibrium states are simply the intersections between the v_R-i_R curve and the load line. To apply the corollary to the stability criteria, we obtain

$$f'(v_C) = \frac{1}{C}\left[-\frac{1}{R} - g'(v_C)\right] \tag{14-5}$$

where $g'(v_C)$ represents the slope of the v_R-i_R curve, and is called the *incremental conductance*. From Eq. (14-5) we conclude that an equilibrium state Q_j is stable if, and only if,

$$\frac{1}{R} > -g'(v_R)|_{Q_j} \tag{14-6}$$

where $-g'(v_R)|_{Q_j}$ is the incremental conductance at the point Q_j. Equation (14-6) represents the stability criterion for the network in Fig. 14-3*a* and can be stated as follows: An equilibrium state

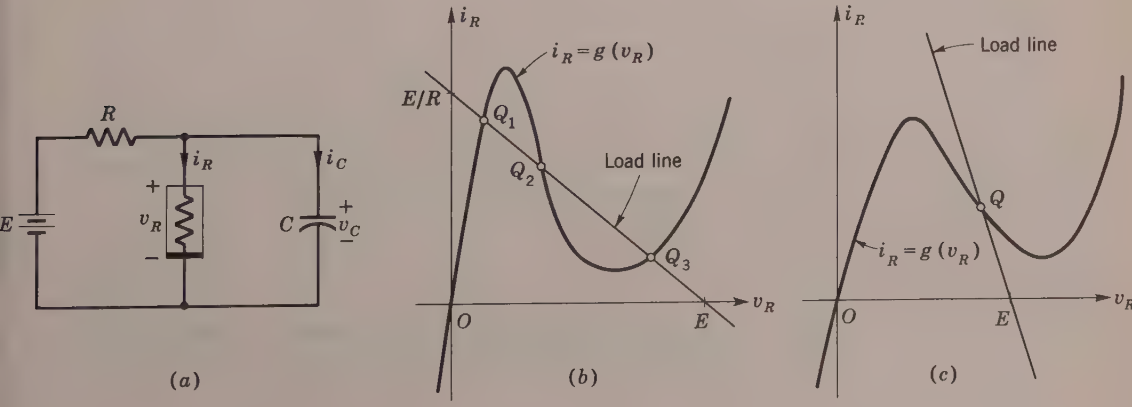


Fig. 14-3. An equilibrium state Q_j of this network (configuration 1) is stable if, and only if, the conductance $1/R$ is greater than the magnitude of the incremental conductance of the v_R-i_R curve at Q_j.

Q_j on the v_R-i_R curve is always stable if the incremental conductance $g'(v_R)$ at Q_j is positive. However, if $g'(v_R) < 0$, then Q_j is stable only if the magnitude of the incremental conductance at Q_j is less than the conductance $1/R$ of the load line.

This criterion enables us to find the stability of the equilibrium state by inspection, i.e., without obtaining the DP plot. For example, suppose the nonlinear resistor is a tunnel diode with the v_R-i_R curve shown in Fig. 14-3*b*. Since the slope of this curve at Q_1 and Q_3 is positive, we conclude immediately that Q_1 and Q_3 are stable. Since the slope at Q_2 is negative, and the magnitude of the incremental conductance $|g'(v_R)|$ is *not* less than $1/R$, we conclude that Q_2 is unstable. On the other hand, the equilibrium state Q in Fig. 14-3*c* is stable because the incremental conductance at Q is less than $1/R$.

Configuration 2 Consider next the network shown in Fig. 14-4*a*. This time it is more convenient to denote the v_R-i_R curve by $v_R = r(i_R)$. The corresponding normal-form equation is given by

$$\frac{di_L}{dt} = \frac{(E - Ri_L) - r(i_L)}{L} \tag{14-7}$$

Comparing this equation with Eq. (14-1), we obtain

$$f(i_L) = \frac{(E - Ri_L) - r(i_L)}{L} \tag{14-8}$$

The equilibrium states are given, as usual, by the intersections between the load line $v = E - Ri_L$ and the v_R-i_R curve. To apply the corollary to the stability criteria, we obtain

$$f'(i_L) = \frac{1}{L}[-R - r'(i_L)] \tag{14-9}$$

where $r'(i_L)$ represents the reciprocal slope of the v_R-i_R curve and is called the *incremental resistance*. From Eq. (14-9) we conclude that an equilibrium state Q_j is stable if, and only if,

$$R > -r'(i_R)|_{Q_j} \tag{14-10}$$

where $-r'(i_R)|_{Q_j}$ is the incremental resistance at the point Q_j. Equation (14-10) represents the stability criterion for the network

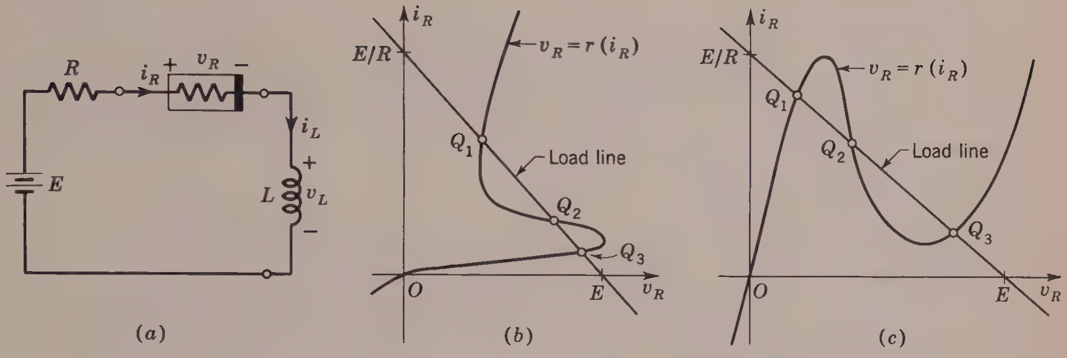


Fig. 14-4. An equilibrium state Q_j of this network (configuration 2) is stable if, and only if, the resistance R is greater than the magnitude of the incremental resistance of the v_R-i_R curve at Q_j.

in Fig. 14-4*a* and can be stated as follows: An equilibrium state Q_j on the v_R-i_R curve is always stable if the incremental resistance $r'(i_R)$ at Q_j is positive. However, if $r'(i_R) < 0$, then Q_j is stable only if the magnitude of the incremental resistance at Q_j is less than the resistance R of the load line.

As an application of this criterion, suppose the nonlinear resistor is a neon bulb with the v_R-i_R curve shown in Fig. 14-4*b*. By inspection we immediately conclude that the equilibrium states Q_1 and Q_3 are stable and that Q_2 is unstable. On the other hand, if we replace the neon bulb by a tunnel diode, as in Fig. 14-4*c*, then we find that all three equilibrium states Q_1, Q_2, and Q_3 are stable. This conclusion appears at first sight to contradict the conclusion of the corollary to the stability criteria, namely, that the equilibrium states must alternate between stability and instability. There is no contradiction, however, because the expression $f(i_L)$ in Eq. (14-8) is *multivalued* for a voltage-controlled curve, and the alternating property is true only if $f(i_L)$ is a single-valued function of $i_L = i_R$, such as the neon bulb curve in Fig. 14-4*b*.[1]

Exercise 1: Sketch the curve $f(v_C)$ given by Eq. (14-4) corresponding to the v_R-i_R curve and the resistance R in Fig. 14-3*b* and *c*. Use this curve to verify the stability criterion for configuration 1.

Exercise 2: Derive the normal-form equations (14-3) and (14-7). Formulate a simple rule so that these equations may be written by inspection.

Exercise 3: Sketch the curve $f(i_L)$ given by Eq. (14-8) corresponding to the v_R-i_R curve and the resistance R in Fig. 14-4*b* and *c*. Use this curve to verify the stability criterion for configuration 2.

Exercise 4: Prove that the stability criteria and their corollary (except the alternating property) are valid even if $f(x)$ is a multivalued function.

---

[1] It is important to recognize that while the tunnel-diode v_R-i_R curve is a multivalued function in configuration 2, it is a single-valued function in configuration 1. Hence, we must always choose the pertinent independent variable before deciding whether a curve is a single-valued function.

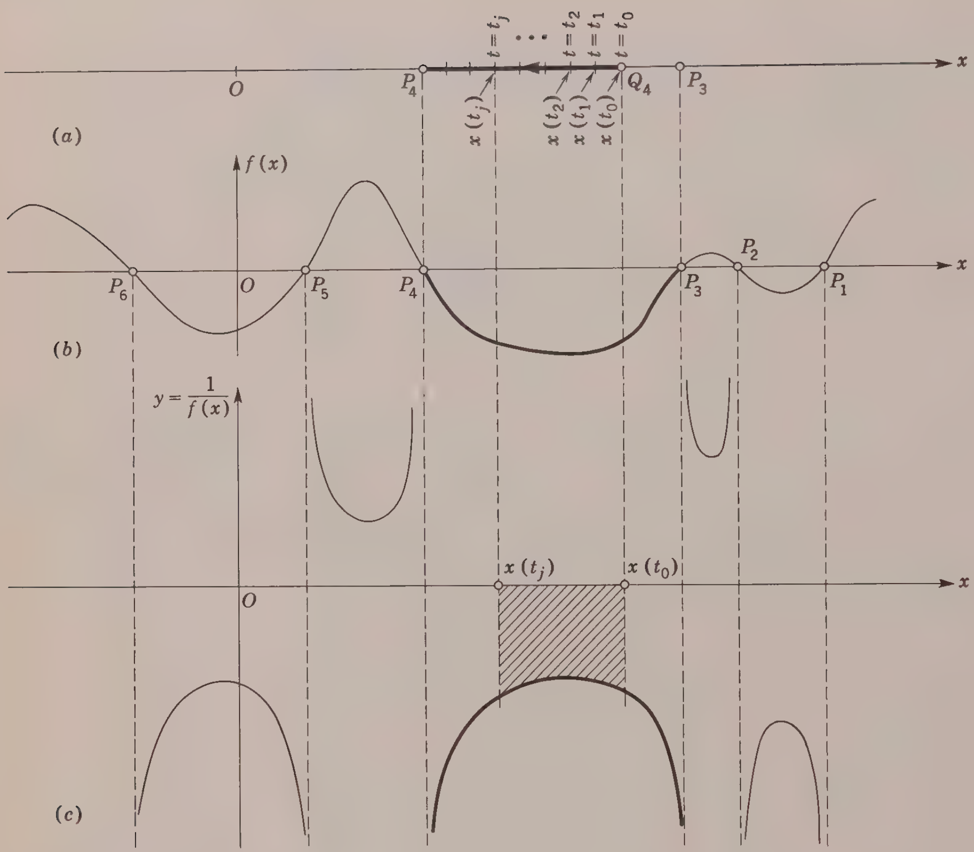


Fig. 14-5. The value of t corresponding to each point on the trajectory is equal to the area under the curve $1/f(x)$ from the initial state to the point in question, plus the initial time t_0.

14-4 TIME SCALING THE TRAJECTORY OF SOLUTIONS

The trajectories we obtained in the preceding section are incomplete in the sense that the time scale associated with each trajectory is absent. We shall now show how this information can be obtained.

Consider a typical trajectory such as the one with the initial state at Q_4 as shown in Fig. 14-2*b*. For convenience let us redraw this trajectory, together with the portion of $f(x)$ associated with it as shown in Fig. 14-5*a* and *b*. The initial state corresponds, of course, to the initial time $t = t_0$. Our problem is to evaluate the value of t associated with any other point on the trajectory. To be specific, let us choose an arbitrary point $x = x(t_j)$ as shown in Fig. 14-5*a* and try to evaluate the corresponding time $t = t_j$. To do this, let us examine Eq. (14-1) again and recast it in the form

$$dt = \frac{1}{f(x)}\,dx \tag{14-11}$$

We can integrate both sides of this equation to obtain

$$\int_{t_0}^{t_j} dt = \int_{x(t_0)}^{x(t_j)} \frac{1}{f(x)}\,dx \tag{14-12}$$

Observe that since the limits of integration on the left side are from t_0 to t_j, the corresponding limits of integration on the right side must be from $x(t_0)$ to $x(t_j)$. Integrating the left side of Eq. (14-12) and simplifying, we obtain

$$t_j = t_0 + \int_{x(t_0)}^{x(t_j)} \frac{1}{f(x)}\,dx \tag{14-13}$$

This equation gives an explicit formula for calculating t_j corresponding to any point $x(t_j)$ on the trajectory. In particular, it shows that t_j is simply equal to the initial time t_0 plus the area under the curve $y = 1/f(x)$ from $x = x(t_0)$ to $x = x(t_j)$ as shown in Fig. 14-5*c*. This area can be obtained by a numerical integration procedure using a computer or a planimeter. It can also be determined by a number of approximate methods such as the trapezoidal rule.

We can now summarize the procedure for finding the time scales as follows:

1. Draw the curve $y = 1/f(x)$ as a function of x over a range coinciding with that of the trajectory.
2. Find the area under the curve $y = 1/f(x)$ from the initial state t_0 to any desired point on the trajectory.
3. The sum of the area obtained in (2) and the value t_0 gives the corresponding time at the desired point.

From the above procedure we can deduce a number of interesting properties concerning the nature of the solution $x = X(t)$. First of all, let us observe that each trajectory must lie either between two equilibrium states or between one equilibrium state and $x = +\infty$ or $x = -\infty$. This means that the portion of the curve $y = 1/f(x)$ associated with each trajectory must lie either above

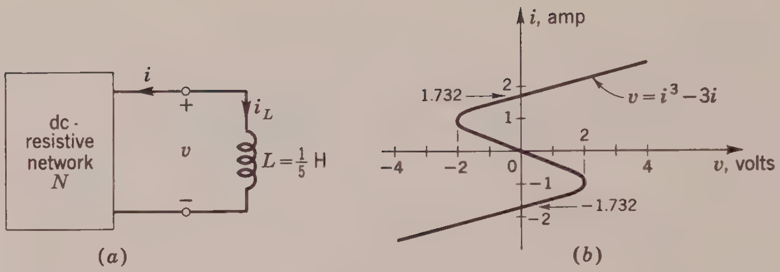


Fig. 14-6. A first-order network consisting of a linear inductor connected across a current-controlled dc-resistive network.

the x axis or below the x axis. Therefore, the area under the curve $y = 1/f(x)$ is always positive and increases as the trajectory moves toward the equilibrium state. This means that the solution $x = X(t)$ must necessarily be a monotonically increasing or a monotonically decreasing function of t.†

Let us observe further that at each equilibrium state, the curve $y = 1/f(x)$ approaches either $+\infty$ or $-\infty$ because $f(x) = 0$ at this point. It can be seen, therefore, that as x approaches an equilibrium state, the area increases rapidly, and more time is needed for the trajectory to move the same distance Δx. This can be interpreted by saying that the motion becomes slower. In fact, except in some ideal cases where the curve $y = 1/f(x)$ is integrable up to the equilibrium state,[1] the area is generally infinite. This implies that, except in some special cases, it will generally take an infinite time interval for the trajectory to reach its equilibrium state. We shall summarize the above observations into a theorem.

MONOTONIC SOLUTION THEOREM

If a first-order network can be described by a normal-form equation $dx/dt = f(x)$, where $f(x)$ is a *single-valued function* of x, then its solution will always be a monotonically increasing or a monotonically decreasing function of x. Moreover, except for some special cases, it will require an *infinite* time interval for the solution to reach its steady-state value.

Let us now illustrate the time-scaling procedure by an example.

EXAMPLE

Consider the network shown in Fig. 14-6a where the DP plot of N is given by the current-controlled curve shown in Fig. 14-6b.[2] If the current i_L at $t = 1$ sec is given to be -0.25 amp, find the inductor current $i_L(t)$ for $t \geq 1$ sec.

---

† Throughout this section we assume that $f(x)$ is a single-valued function of x, and our conclusion is valid only under this condition.

[1] In mathematics the integral of $y = 1/f(x)$ is said to be improper if $y = \pm\infty$ at the lower or upper limits of integration. In some special cases, the area can be found by a limiting operation. If the answer turns out to be a finite number, the curve is said to be integrable.

[2] For convenience we have chosen the simple curve $v = r(i) = i^3 - 3i$. In practice the DP plot does not usually have an analytical expression and must be specified graphically.

Since the DP plot is specified in the i-vs.-v plane, it is usually more convenient to choose i instead of i_L as the state variable; after all, i and i_L differ only by a negative sign. The normal-form equation is given by

$$\frac{di}{dt} = -5r(i) = -5(i^3 - 3i) \tag{14-14}$$

From this we identify the functions $f(x)$ and $1/f(x)$ as

$$f(i) = -5i^3 + 15i \tag{14-15}$$

$$\frac{1}{f(i)} = \frac{1}{-5i^3 + 15i} \tag{14-16}$$

The equilibrium states of this network are obtained by plotting the function $f(i)$ as shown in Fig. 14-7a and determining the points of intersection with the i axis, namely, $Q_1(i = -\sqrt{3})$, $Q_2(i = 0)$, and $Q_3(i = \sqrt{3})$. Applying the corollary to the stability criteria,

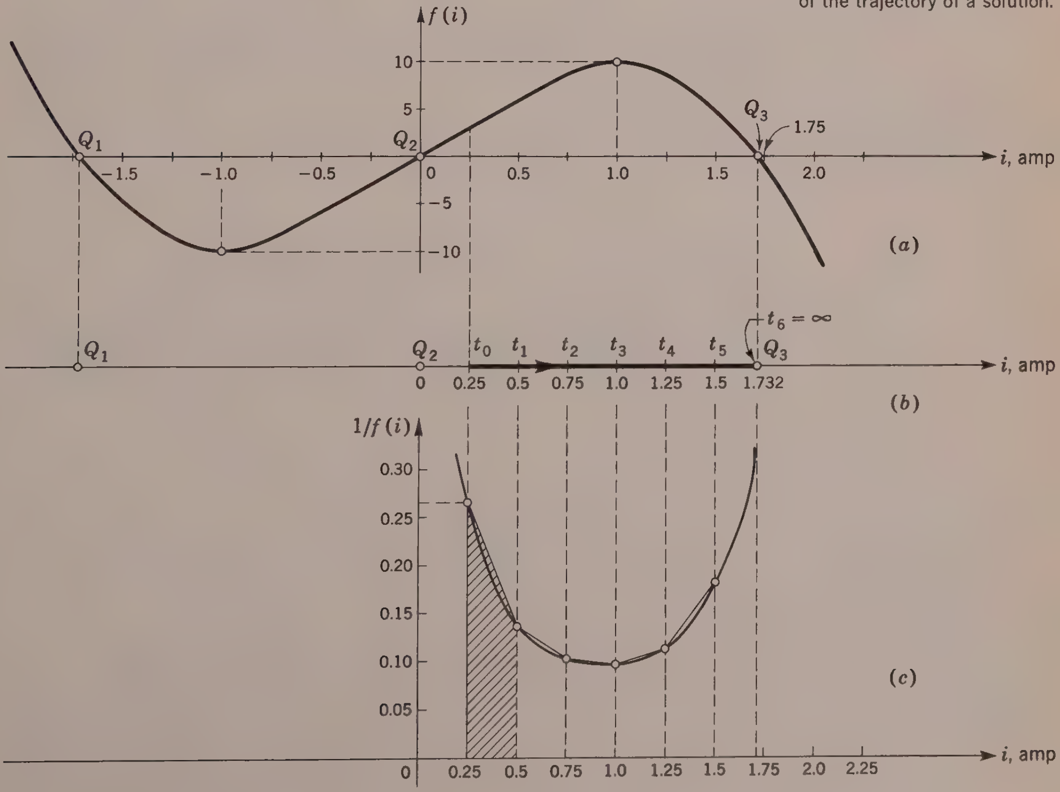
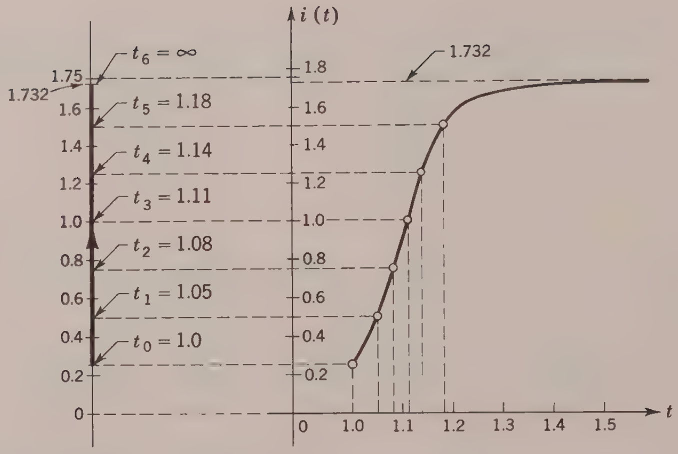


Fig. 14-7. Graphical procedure for determining the time scale of the trajectory of a solution.

$i(t_j)$, amp	t_j, sec
$i(t_0) = 0.25$	$t_0 = 1.0$
$i(t_1) = 0.5$	$t_1 = 1.05$
$i(t_2) = 0.75$	$t_2 = 1.08$
$i(t_3) = 1.0$	$t_3 = 1.11$
$i(t_4) = 1.25$	$t_4 = 1.14$
$i(t_5) = 1.5$	$t_5 = 1.18$
$i(t_6) = 1.732$	$t_6 = \infty$

(*a*)



(*b*) (*c*)

Fig. 14-8. The solution waveform of $i(t)$ in (*c*) is obtained from the time-scaled trajectory in (*b*).

we conclude that Q_1 and Q_3 are stable but Q_2 is unstable. Therefore, corresponding to the given initial state $i(t_0 = 1) = 0.25$ amp, the trajectory shown in Fig. 14-7*b* is obtained. Finally, to determine the time scale associated with this trajectory, we plot the function $1/f(i)$ as shown in Fig. 14-7*c* and apply Eq. (14-13) to obtain

$$t_j = 1 + \int_{0.25}^{i(t_j)} \frac{1}{-5i^3 + 15i}\, di \tag{14-17}$$

For this example we have subdivided the trajectory of Fig. 14-7*b* into six equal divisions. The corresponding value of t_j at each division can be obtained by either integrating Eq. (14-17) directly or by determining the shaded area under the curve as shown in Fig. 14-7*c*. We have chosen the latter by applying the trapezoidal rule.[1]

$$t_j = t_{j-1} + \frac{1}{2}\left\{\frac{1}{f[i(t_{j-1})]} + \frac{1}{f[i(t_j)]}\right\}[i(t_j) - i(t_{j-1})]$$

For example, the value of t_4 can be approximately calculated as follows:

$$\begin{aligned} t_4 &= 1.11 + \frac{1}{2}\left[\frac{1}{-5(1.0)^3 + 15(1.0)} + \frac{1}{-5(1.25)^3 + 15(1.25)}\right](1.25 - 1.0) \\ &= 1.14 \text{ sec} \end{aligned} \tag{14-18}$$

---

[1] This consists of approximating the area under the curve by a number of trapezoids. The resulting answer is, of course, not exact. But by choosing enough subdivisions, the results can be obtained to within any prescribed accuracy.

The results are tabulated as shown in Fig. 14-8*a*, and the complete trajectory is plotted in Fig. 14-8*b*. From this we obtain the time function $i(t)$ as shown in Fig. 14-8*c*. The final solution $i_L(t)$ is then simply the negative of $i(t)$. The reader should recognize that this example is given mainly for illustrative purposes. In practice, a digital computer could obviously be used to compute the time scales automatically.

14-5 REALISTIC AND INCOMPLETE MODELS

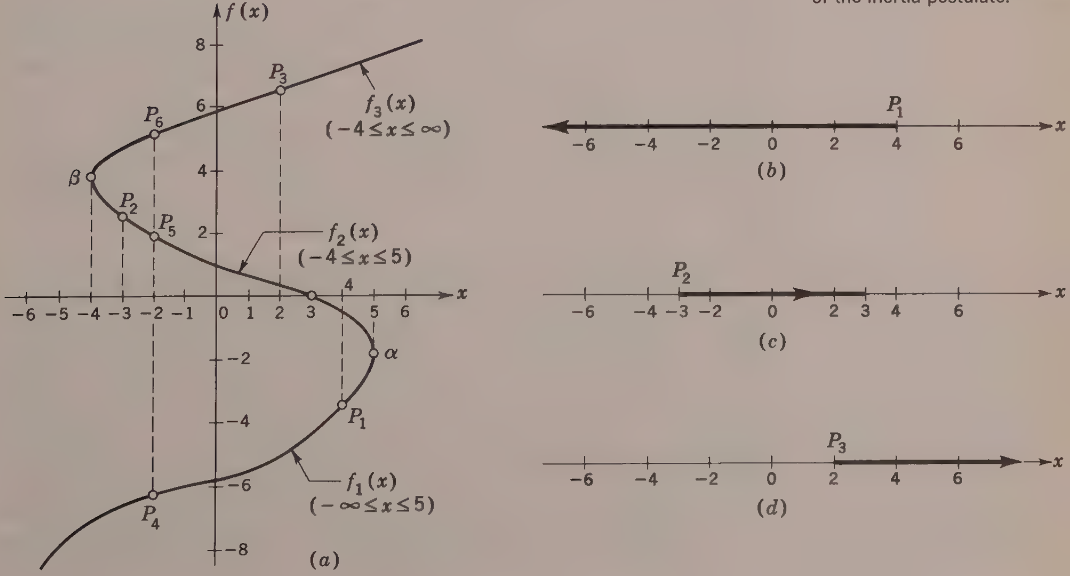
So far we have tacitly assumed that $f(x)$ is a single-valued function. Under this assumption we have shown that the solution $x(t)$ of a first-order autonomous network is either a monotonically increasing or a monotonically decreasing function of time. We shall now investigate the nature of the solution when $f(x)$ is a multivalued function of x, such as the curve shown in Fig. 14-9*a*.

Mathematically, the normal-form equation

$$\frac{dx}{dt} = f(x)$$

is meaningless when $f(x)$ is multivalued because this implies that the solution $x(t)$ may possess more than one slope at a given time. However, we know that any dynamic network in practice must

Fig. 14-9. An example illustrating how the inconsistency arising from a multivalued function can be resolved with the help of the inertia postulate.



have a unique solution for all times. This paradox can be resolved by a clear understanding of the differences between a theoretical result and an experimental result. When we try to analyze a network on paper, we are really analyzing a model of the network and not the actual network. As we have emphasized many times, a model is never an exact representation of a physical network. If the solution obtained from using a certain model agrees reasonably well with actual measurements, the model is said to be *realistic*. On the other hand, if a model leads to an inaccurate solution, it is said to be *unrealistic*. In the extreme case where an unrealistic model leads to physical inconsistencies, such as the multivalued example above, then it is said to be *incomplete.* As will be seen in the sequel, any inconsistency arising from the use of an incomplete model can usually be resolved by introducing one or more parasitic inductances or capacitances at some strategic locations in the network. In other words, an incomplete model usually results when certain parasitic elements cannot be neglected.

There are two methods for handling incomplete models. The most general method is to refine the model so that it becomes more realistic. Of course, the more parasitic elements we include, the better will be the model. However, not all parasitic elements play a major role. In fact, it has been found from experience that most parasitic elements can be neglected without introducing significant error. Therefore, in order to obtain a *simple* yet realistic model, we should include only those parasitic energy-storage elements which play a dominant role. These elements will be called *major energy-storage elements.* Unfortunately, it is not always easy to find these elements in a given model.[1]

A less general but more practical method to resolve an inconsistency arising from an incomplete model is to formulate a *postulate* based upon experimental observations.[2] We shall now discuss a postulate which can be used to handle many incomplete models.

Any multivalued function $f(x)$ can be broken up into two or more single-valued functions $f_1(x)$, $f_2(x)$, . . . , $f_n(x)$ where each function $f_j(x)$ is defined over an interval $a_j \leq x \leq b_j$. For example, the multivalued function $f(x)$ shown in Fig. 14-9*a* can be broken up into three single-valued functions $f_1(x)$, $f_2(x)$, and $f_3(x)$. The function $f_1(x)$ consists of the lower branch below point α, and is defined over the interval $-\infty \leq x \leq 5$. The function $f_2(x)$ consists of the middle branch between points α and β, and is defined over the interval $-4 \leq x \leq 5$. The function $f_3(x)$ consists of the upper branch above point β and is defined over the interval $-4 \leq x \leq \infty$. From experimental evidence it has been observed that whenever the solution $x(t)$ is operating initially on a given branch of $f(x)$, it will continue to operate on that branch until $x(t)$ moves outside the interval of definition of that branch. This observation is analogous to inertia in physical systems. For example, if the initial state is given to be at point P_1 of the lower branch $f_1(x)$, as shown in Fig. 14-9*a*, the network will tend to see only the branch $f_1(x)$, and the trajectory of the solution can be obtained by solving the normal-form equation with $f(x)$ replaced by $f_1(x)$. Observe that the above procedure is based only on the argument that the solution agrees with experimental observations. Mathematically, there is no valid reason to assume that this is what happens. For example, it is perfectly valid to raise the following objection: Why should not the solution starting from P_1 on $f_1(x)$ (Fig. 14-9*a*) arrive at some point such as P_4 on $f_1(x)$ and suddenly jump to P_5 on $f_2(x)$ or P_6 on $f_3(x)$? After all, for each value of x, there are three possible operating points! While it is not possible to dispel this objection, it is equally valid to counter it with the question: Why should this happen? The only way to resolve the issue is to state our assertion in the form of a postulate.

---

[1] The problem of determining the major energy-storage elements associated with a model is still a current research topic.

[2] A postulate is a rule whose validity is not contradicted by experimental observations; thus, it is assumed to be true. It is not possible to prove a postulate.

INERTIA POSTULATE

Let $f(x)$ be a multivalued function consisting of n single-valued branches $f_1(x), f_2(x), \ldots, f_n(x)$, where each branch $f_j(x)$ is defined over an interval $a_j \leq x \leq b_j$. Let the initial state $x(t_0)$ be given as a point on one of these branches, say $f_k(x)$. Then the network can be assumed to remain operating on this branch, and the normal-form equation can be modified by replacing the multivalued function $f(x)$ by the single-valued function $f_k(x)$. The resulting solution $x(t)$ is a valid solution for all times $t \geq t_0$ for which $a_k \leq x(t) \leq b_k$.

To illustrate the application of the inertia postulate, let us apply it to the multivalued function $f(x)$ of Fig. 14-9*a* again. Suppose the initial state is given to be at P_1. We can conclude that the pertinent branch is $f_1(x)$, and the normal-form equation is

$$\frac{dx}{dt} = f_1(x) \qquad -\infty \leq x \leq 5 \tag{14-19}$$

The resulting trajectory of solution is shown in Fig. 14-9*b*.

If the initial state is given to be at P_2, then the pertinent branch is $f_2(x)$, and the normal-form equation is

$$\frac{dx}{dt} = f_2(x) \qquad -4 \leq x \leq 5 \tag{14-20}$$

The resulting trajectory is shown in Fig. 14-9*c*. Observe that the terminal point $x = 3$ is an equilibrium state.

Finally, if the initial state is given to be at P_3, then the pertinent branch is $f_3(x)$, and the normal-form equation is

$$\frac{dx}{dt} = f_3(x) \qquad -4 \leq x \leq \infty \tag{14-21}$$

The resulting trajectory is shown in Fig. 14-9*d*.

14-6 JUMP PHENOMENON AND OSCILLATORY SOLUTION IN FIRST-ORDER NETWORKS

In the multivalued example shown in Fig. 14-9, we were able to find the solution $x(t)$ for all times $t \geq t_0$ by applying the inertia postulate. We shall now demonstrate a new phenomenon which requires additional principles in order to obtain the solution for all times $t \geq t_0$.

Consider the multivalued function $f(x)$ shown in Fig. 14-10*a*. Applying the inertia postulate to the various initial states $P_1(5,-15)$, $P_2(-15,-5)$, $P_3(5,5)$, and $P_4(-15,20)$, we obtain the corresponding trajectories shown in Fig. 14-10*b* to *e*, respectively. Observe, however, that unlike the examples in the preceding section, each of these trajectories moves in such a direction that it tends to leave the branch of $f(x)$ originally containing the initial state. For example, the trajectory starting at P_1 on the lower branch $f_1(x)$ must move toward the left, thus causing the operating point to move along the curve toward point α as shown in Fig. 14-10*a*. But this point is not an equilibrium state. Therefore, if we determine the time scale of the trajectory of Fig. 14-10*b*, the point $\alpha(x = -20)$ will be reached after some *finite* time interval, say at $t = t_j$. The question now is, What happens at $t > t_j$? One might be tempted to guess that the trajectory, upon reaching point α, might move up along the middle branch $f_2(x)$. But this is impossible, for at point α we have $f(x) = -10 < 0$; hence, $dx/dt < 0$. This means that the value of x must continue to decrease, a condition contradictory to our initial guess. Therefore the operating point cannot continue on to the middle branch; neither can it turn

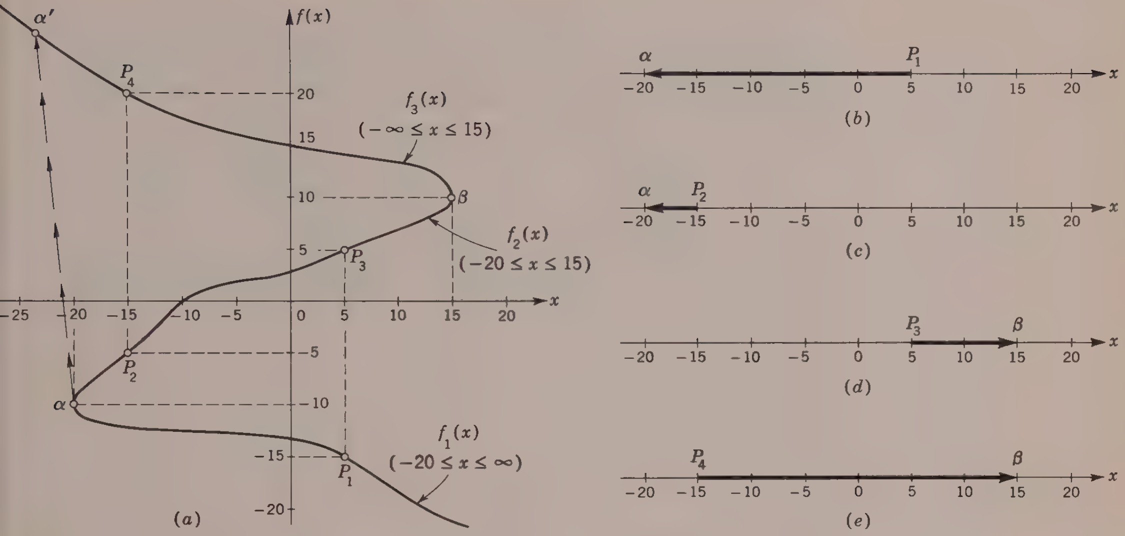


Fig. 14-10. An example illustrating the mechanism that leads to a jumping phenomenon.

around and go back on the lower branch. Where can it go then? Certainly, the inertia postulate is useless here because the trajectory has moved outside the interval of definition of $f_1(x)$. If we examine the remaining three trajectories shown in Fig. 14-10*c* to *e*, we would in each case have arrived at the same difficulty. This dilemma cannot be resolved mathematically, for as we have already stated, the normal-form equation is meaningless whenever $f(x)$ is multivalued.

Since the trajectory, upon reaching point α, can neither continue to $f_2(x)$ nor stay on $f_1(x)$, it must *jump* to some other point α' (such as the point shown in Fig. 14-10*a*) on the curve $f(x)$ and this *transition* must occur instantaneously, for at any time the operating point must stay on $f(x)$ in order to qualify as a solution. We shall call the operating point α the *terminal state* and the new operating point α' the *transition state* of the trajectory.

Since the transition from α to α' is instantaneous, the time $t = t_j$ when the network is operating at α must also be the time when the network jumps to α'. However, it is convenient for us to split the time $t = t_j$ into $t = t_j^-$ and $t = t_j^+$ and associate the operating point α before the jump with t_j^- and the operating point α' after the jump with t_j^+. For example, corresponding to point α of Fig. 14-10*a*, we have

$$x(t_j^-) = x_\alpha = -20$$

and

$$x(t_j^+) = x_{\alpha'}$$

If we can find $x(t_j^+)$, then we can apply the inertia postulate once again to determine the continuation of the solution for $t > t_j^+$. Our crucial problem, therefore, depends on finding the value $x(t_j^+)$ of the transition state α'.

Let us recall that the state variable x may represent either the charge q or voltage v_C across a capacitor, or the flux linkage φ or current i_L in an inductor. To be specific, suppose q is the state variable and our problem is, given $q = q(t_j^-)$ at a terminal state (such as α), find $q = q(t_j^+)$ at the corresponding transition state. Since

$$
\begin{aligned}
q(t_j^+) &= \int_{-\infty}^{t_j^+} i(\tau)\,d\tau \\
&= \int_{-\infty}^{t_j^-} i(\tau)\,d\tau + \int_{t_j^-}^{t_j^+} i(\tau)\,d\tau \qquad (14\text{-}22)
\end{aligned}
$$

and since

$$\int_{-\infty}^{t_j^-} i(\tau)\,d\tau = q(t_j^-)$$

Eq. (14-22) can be written as

$$q(t_j^+) = q(t_j^-) + \int_{t_j^-}^{t_j^+} i(\tau)\,d\tau \qquad (14\text{-}23)$$

Observe that the integral term in Eq. (14-23) represents the area under a curve of zero arc length and is, therefore, zero unless $i(t_j^-)$ happens to be infinity. But no physical device can supply an infinite current; therefore this integral is zero, and Eq. (14-23) reduces to

$$q(t_j^+) = q(t_j^-) \qquad (14\text{-}24)$$

This equation tells us that the value of q must remain the same during the transition. In other words, we have shown that the charge across a capacitor cannot change instantaneously and hence $q(t)$ must be a continuous function of time. This important property is often referred to in the literature as the *principle of conservation of charge.*

In terms of the example shown in Fig. 14-10*a*, this principle allows us to conclude that if x represents the capacitor charge q, then when the trajectory reaches the terminal state α as shown in Fig. 14-11*a* and *b* at $t = t_j^-$, the transition state α' must lie directly above α. In other words, the transition must take place instantaneously along the vertical path shown in Fig. 14-11*a*. Once we find α', we can apply the inertia postulate again and draw the trajectory as shown in Fig. 14-11*c*. Observe that as soon as this trajectory reaches its terminal state $\beta(q = 15)$ at $t = t_k^-$, it will jump instantaneously along the vertical path shown in Fig. 14-11*a* to the transition state β'. This will allow us to apply the inertia postulate once again, and the operation will clearly repeat itself periodically. If we plot the resulting solution of $q(t)$ as a function of time, we obtain the oscillatory waveform shown in Fig. 14-11*d*. We shall show later that this mechanism is fundamental in the design of oscillators.

The above example was based on the assumption that the state variable x represents the capacitor charge q so that we can apply the principle of conservation of charge. If the state variable x represents the capacitor voltage v_C, we can invoke the v_C-q curve of the capacitor to conclude that

$$v_C(t_j^-) = v_C(t_j^+) \tag{14-25}$$

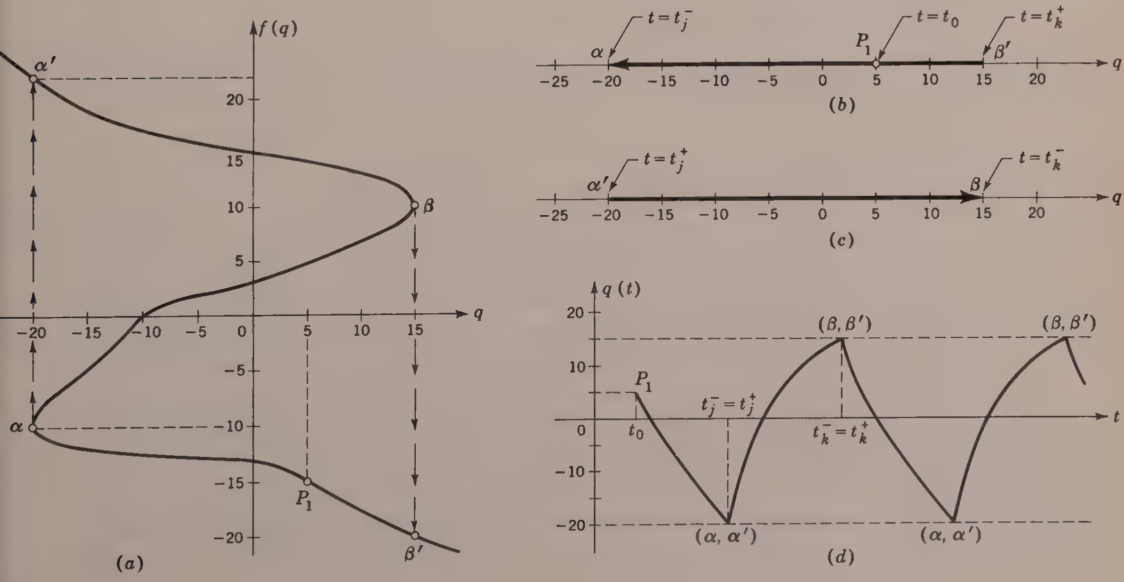


Fig. 14-11. An illustration showing how the jumping phenomenon can lead to an oscillatory solution waveform.

By an exact dual procedure we can conclude that in case the state variable x represents either the flux linkage φ or the current i_L of an inductor, we have[1]

$$\varphi(t_j^-) = \varphi(t_j^+) \tag{14-26}$$

or

$$i_L(t_j^-) = i_L(t_j^+) \tag{14-27}$$

This property is often referred to in the literature as the *principle of conservation of flux linkages.*

As a summary, we have shown that whenever the trajectory reaches a terminal state representing the end point of a multivalued branch, it must jump instantaneously to another point on the multivalued curve $f(x)$ called the *transition state.* The exact location of this transition state can be determined by invoking the principle of conservation of charge (flux linkages) if the state variable x represents q or v_C (φ or i_L) of a capacitor (inductor). In view of their importance, we shall restate these principles as follows:

PRINCIPLE OF CONSERVATION OF CHARGE

In the absence of an infinite current at $t = t_j^-$, the charge q and the voltage v_C across a capacitor cannot change instantaneously.

PRINCIPLE OF CONSERVATION OF FLUX LINKAGES

In the absence of an infinite voltage at $t = t_j^-$, the flux linkage φ and the current i_L in an inductor cannot change instantaneously.

14-7 INCOMPLETE MODELS REQUIRING ADDITIONAL PARASITIC ELEMENTS

In Sec. 14-5 we classified any network model which leads to physical inconsistencies as being incomplete. This class of incomplete network models arises as a result of an oversimplification in the components of the corresponding physical networks, namely, whenever certain major parasitic capacitors or inductors cannot be neglected. The only valid method to resolve this inconsistency is to include the major energy-storage elements, thereby increasing the order of complexity of the resulting networks. Since higher-order networks are much harder to analyze, we tried to preserve the simplicity of first-order networks by introducing the inertia postulate and the principles of conservation of charge and flux linkages.

---

[1] This property is seen to be true in all cases except when the voltage across the inductor is infinite, a condition unattainable in practice.

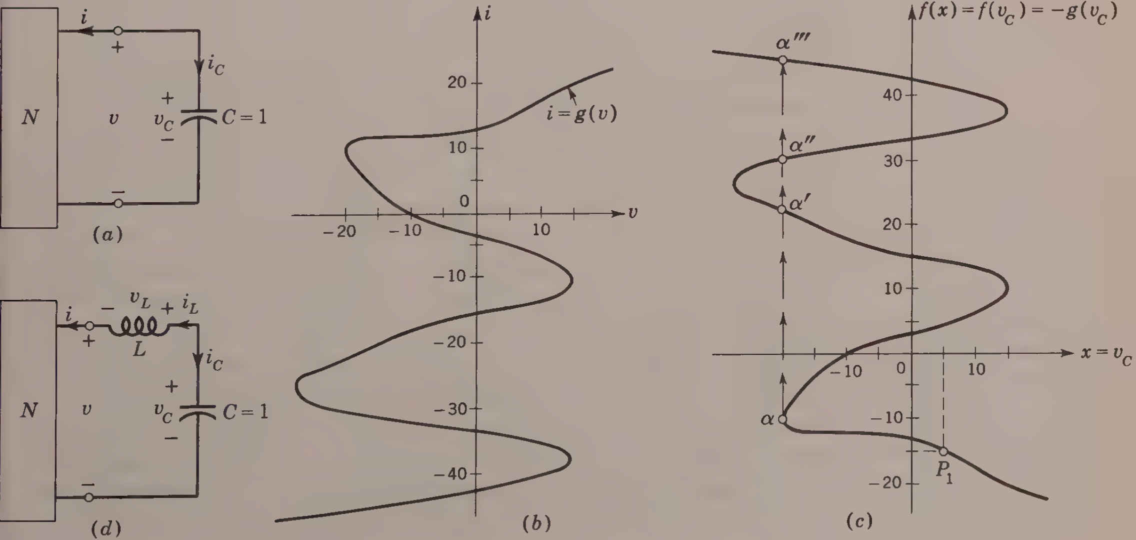
These "gimmicks" will allow us to analyze a large class of practical dynamic networks as first-order networks. However, the *only* justification for introducing these gimmicks is that they allow us to resolve the inconsistency successfully. There is no guarantee that these gimmicks will work in all cases. In fact, we shall now show that there are occasions where it is necessary to include additional parasitic elements.

One occasion to which these gimmicks do not apply occurs when the multivalued function $f(x)$ has more than three branches. For example, consider the network shown in Fig. 14-12*a*, where the DP plot of the black box is shown in Fig. 14-12*b*. The normal-form equation is given by

$$\frac{dv_C}{dt} = -i = -g(v_C) \tag{14-28}$$

In this case the state variable x is v_C, and the function $f(x)$ is given by $-g(v_C)$, as shown in Fig. 14-12*c*. Observe that the lower part of this multivalued curve is identical with that shown in Fig. 14-10*a*. Therefore, if we assume the same initial state $P_1(v_C = 5\text{ volts})$ as before, the operating locus will move from P_1 toward the terminal state α. Upon reaching α, the locus must jump upward vertically to a transition state on the curve. However, there are now three possible candidates, namely, α', α'', or α'''. There is no physical principle that will allow us to decide which of these

Fig. 14-12. An example showing how a multivalued function can lead to nonunique solutions.



points is actually the transition state. Hence, if we insist on using this incomplete model, we must assume each of the three transition states as equally likely and obtain a solution in each case. The result is that we have three distinct solutions, whereas only one of them will actually be observed in practice.

We shall now show that the above difficulty can be easily resolved by introducing a parasitic inductance L in series with the capacitor as shown in Fig. 14-12*d*. This parasitic element may represent the line inductance of the wire connecting the capacitor to the black box. Let us write the normal-form equation of the resulting second-order network in terms of the state variables i_L and v_C. Since $i = i_L$, the v-i curve $v = r(i)$ shown in Fig. 14-12*b* can also be written as $v = r(i_L)$. The resulting normal-form equations are given by

$$\frac{di_L}{dt} = \frac{1}{L}[v_C - r(i_L)] \tag{14-29}$$

$$\frac{dv_C}{dt} = -\frac{1}{C}\, i_L \tag{14-30}$$

Since $r(i_L)$ is a single-valued function of i_L, these differential equations are well-defined mathematically, and can be solved by the methods to be presented in Chap. 19. If we compare these equations with Eq. (14-28), we shall see how the introduction of the parasitic inductor resolves the inconsistency; namely, the multivalued function $i = g(v)$ is obviated by the introduction of $i = i_L$ as the independent variable, thereby resulting in using the inverse function $v = r(i)$.

Another occasion where additional parasitic energy-storage elements may be required to resolve the inconsistency arises whenever a transition state does not exist. For example, suppose the v-i curve shown in Fig. 14-12*b* is replaced by the multivalued function shown in Fig. 14-13*a*. The corresponding function $f(x)$ is shown in Fig. 14-13*b*. If we choose P_1 again as the initial state, the operating locus will once again reach the terminal state α in a finite time. Since the capacitor voltage v_C cannot change instantaneously, there is no point on the curve $f(x)$ that can qualify as a transition state. Hence there can be no solution in this case. This difficulty is even worse than the case presented earlier, where at least one of the three solutions is valid. One way to resolve this difficulty is to introduce a parasitic inductance as before, thereby obtaining Eqs. (14-29) and (14-30) as the normal-form equations. Another way to resolve this difficulty is to recognize the fact that the v-i curve of

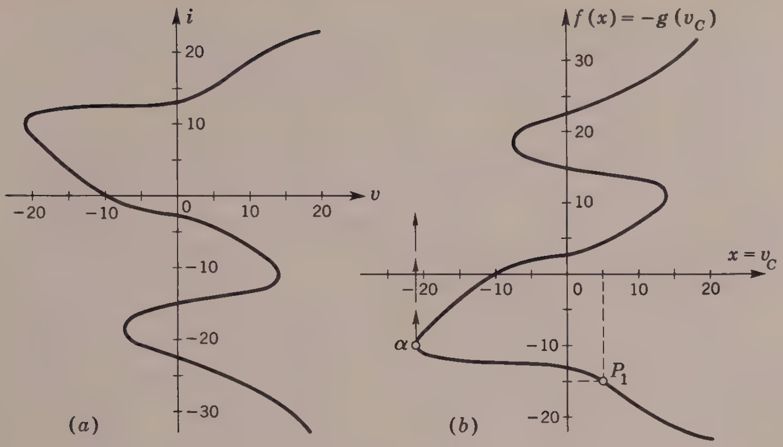


Fig. 14-13. An example illustrating the nonexistence of a transition state.

Fig. 14-13*a* is really *incomplete* in the sense that it possesses some missing branches which are not given. This conclusion is based on the reasoning that no physical network can exhibit a DP plot with a branch that tends to infinity in the fourth quadrant.[1] Invariably, the DP plot in the fourth quadrant must bend toward the third quadrant again, so that it will be in the third quadrant as v tends to $-\infty$. In this case, the corresponding curve $f(x) = -g(v_C)$ in Fig. 14-13*b* will have a branch that tends to infinity in the second quadrant, and a transition state can always be found.

There are more complicated examples where more than one parasitic element may be required, namely, when the DP plot is neither voltage-controlled nor current-controlled. The methods to handle these more difficult cases are beyond the scope of this book. However, it is comfortable to know that most of the common, practical, first-order networks that exist today can be successfully analyzed by the techniques introduced in the preceding sections.

14-8 THE PIECEWISE-LINEAR APPROACH

The method presented in the preceding sections is valuable for obtaining qualitative properties of the network. It is not a very convenient method for obtaining the exact quantitative waveform because of the need to time-scale the trajectories, unless a computer is used. We shall now present a very practical method for obtaining the exact solution waveforms *by inspection.* The only restriction is that all nonlinear elements must be represented by piecewise-linear characteristics. For pedagogical reasons we shall develop this method inductively.

---

[1] Since the power dissipated at any point of the DP plot in the second or the fourth quadrant is negative, the corresponding network is actually *generating* or delivering power to its external load whenever its operating point lies in the second or the fourth quadrant. Hence, an infinite branch in these quadrants would imply an infinite source of power. Review the significance of passivity in Sec. 1-9.

14-8-1 NETWORKS CONTAINING A LINEAR ENERGY-STORAGE ELEMENT ACROSS A NONLINEAR RESISTIVE NETWORK CHARACTERIZED BY PIECEWISE-LINEAR SEGMENTS WITH FINITE NONZERO SLOPES

At any time $t = t_0$, a piecewise-linear network must operate on some linear segment j of the DP plot as seen by the energy-storage element. In view of the inertia postulate the operation will continue to stay on the same segment j until it reaches the breakpoint of this segment at some time $t = t_1$. During the time interval (t_0,t_1), the nonlinear resistive network appears as a linear network to the energy-storage element. Therefore, it can be replaced either by its iterative Thévenin or Norton equivalent circuit as shown in Fig. 14-14*a* and *b* when the energy-storage element is a capacitor, and as shown in Fig. 14-14*c* and *d* when the energy-storage element is an inductor.[1] The corresponding normal-form equation is given by[2]

$$\frac{dv}{dt} = -\frac{1}{r_jC}v + \frac{1}{r_jC}E_j \tag{14-31}$$

$$\frac{dv}{dt} = -\frac{1}{r_jC}v + \frac{1}{r_jC}(-I_jr_j) \tag{14-32}$$

$$\frac{di}{dt} = -\frac{r_j}{L}i + \frac{r_j}{L}\left(-\frac{E_j}{r_j}\right) \tag{14-33}$$

$$\frac{di}{dt} = -\frac{r_j}{L}i + \frac{r_j}{L}I_j \tag{14-34}$$

Observe that these equations are of the same form as

$$\frac{dx}{dt} = -\frac{1}{\tau_j}x + \frac{1}{\tau_j}x_j(t_e) \tag{14-35}$$

where

$$\tau_j = r_jC \tag{14-36}$$

$$x_j(t_e) = E_j = -I_jr_j \tag{14-37}$$

for the capacitor, and

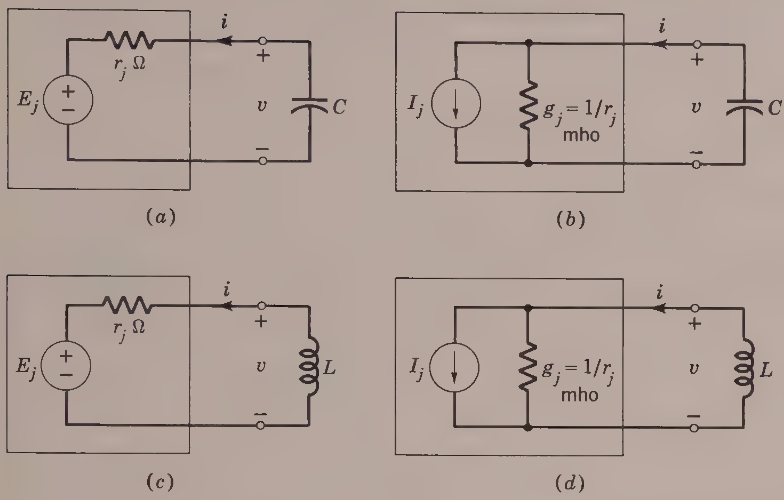
$$\tau_j = \frac{L}{r_j} \tag{14-38}$$

$$x_j(t_e) = I_j = -\frac{E_j}{r_j} \tag{14-39}$$

---

[1] Observe that since only one nonlinear DP plot is involved here, we have chosen the single-subscript symbols r_j, E_j, g_j, and I_j in favor of the double-subscript symbols ${}_kr_j$, ${}_kE_j$, ${}_kg_j$, and ${}_kI_j$ used in Chap. 12.

[2] For reasons that will soon be obvious, it is more convenient to choose i instead of i_L as the state variable.



(a) (b) (c) (d)

Fig. 14-14. During the time interval (t_0,t_1), a first-order nonlinear network is equivalent to a first-order linear network.

for the inductor. The constant τ_j will play an important role later, and will be called the *time constant* corresponding to the segment j of the network. If we equate the right-hand side of Eq. (14-35) to zero, we find that the constant $x_j(t_e)$ is simply the equilibrium state corresponding to segment j of the network. In particular, the equilibrium state defined by Eq. (14-37) is easily seen to be the intersection of segment j (or its extension) with the v axis.[1] Similarly, the equilibrium state defined by Eq. (14-39) is easily seen to be the intersection of segment j (or its extension) with the i axis. It is important to observe that since each segment j of the DP plot is defined over some interval, the value $x_j(t_e)$ at the equilibrium state may fall outside the interval of definition for segment j. In this case we shall call $x_j(t_e)$ a *virtual equilibrium state.* Clearly, the solution $x(t)$ will never reach a virtual equilibrium state because as soon as $x(t)$ reaches the breakpoint of segment j, it will enter a new segment. We shall see shortly, however, that a virtual equilibrium state is useful in determining the solution of a network by inspection.

Basic form of solution Let us now investigate the nature of the solution to Eq. (14-35) corresponding to an initial state $x = x_j(t_0)$ located in segment j. By a direct substitution, we can easily verify that the solution is given by

$$x(t) = x_j(t_e) + [x_j(t_0) - x_j(t_e)]e^{-[(t-t_0)/\tau_j]} \qquad t \geq t_0 \qquad (14\text{-}40)$$

---

[1] Refer to Fig. 12-1 for this interpretation.

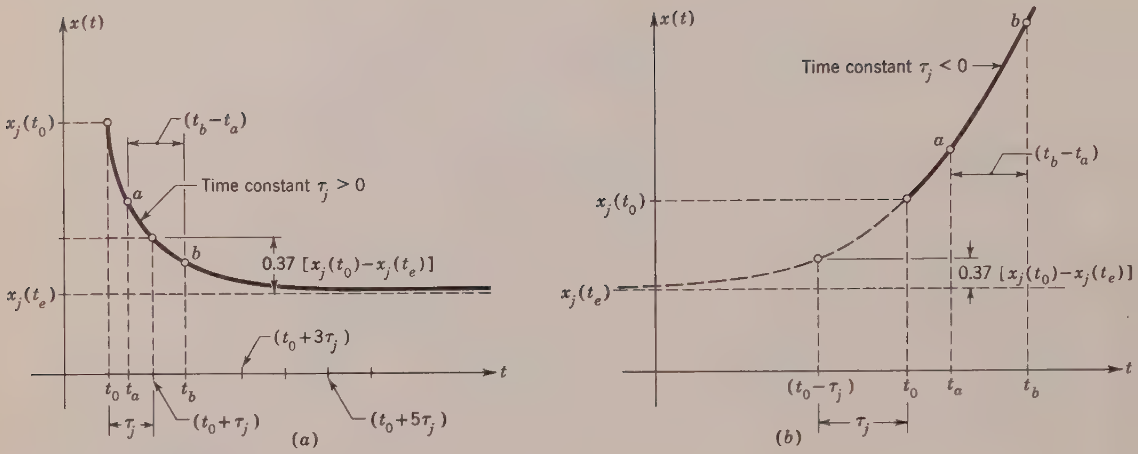


Fig. 14-15. The important parameters associated with an exponential waveform are the initial state $x_j(t_0)$, the equilibrium state $x_j(t_e)$, and the time constant τ_j.

This solution is clearly valid for all times $t \geq t_0$ for which the value $x(t)$ falls within the interval of definition of segment j. An examination of Eq. (14-40) shows that it is completely specified by three parameters, namely,

1. The initial state $x_j(t_0)$
2. The equilibrium state $x_j(t_e)$
3. The time constant τ_j

Given these three parameters, the solution $x(t)$ can be obtained without having to solve the associated differential equation. This observation is the key to our method. Before we discuss this new method, however, it is instructive that we interpret Eq. (14-40) geometrically.

Depending on the sign of the time constant τ_j, Eq. (14-40) represents either a decaying exponential as shown in Fig. 14-15*a* or a growing exponential as shown in Fig. 14-15*b*. Observe that for $\tau_j > 0$, the exponential waveform $x(t) \rightarrow x_j(t_e)$ as $t \rightarrow \infty$.† On the other hand, for $\tau_j < 0$, the exponential waveform $x(t) \rightarrow x_j(t_e)$ as $t \rightarrow -\infty$. Observe further that in the first case, where $\tau_j > 0$, the value of $x(t)$ after one time constant $(t = t_0 + \tau_j)$ is equal to the equilibrium state plus 37 percent of the distance between the initial state and the equilibrium state. In the second case, where $\tau_j < 0$, the value of $x(t)$ at one time constant in the past $(t = t_0 - \tau_j)$ is also equal to the equilibrium state plus 37 percent of the distance between the initial state and the equilibrium state. These observations follow from the fact that

---

† For this reason the equilibrium state corresponding to a positive time constant is often called the *final value.*

$$x(t \pm |\tau_j|) = x_j(t_e) + 0.37[x_j(t_0) - x_j(t_e)] \tag{14-41}$$

where the plus sign corresponds to $\tau_j > 0$, and the negative sign corresponds to $\tau_j < 0$. In view of these properties, we can conclude that when $\tau_j > 0$, the exponential waveform would approach its equilibrium value after a time interval equal to five or more time constants. In other words,

$$x(t) \approx x_j(t_e) \qquad \text{for } t \geq t_0 + 5\tau_j \tag{14-42}$$

Similarly, we can conclude that when $\tau_j < 0$,

$$x(t) \approx x_j(t_e) \qquad \text{for } t \leq t_0 - 5|\tau_j| \tag{14-43}$$

The above results allow us to sketch the waveforms very accurately, knowing only $x_j(t_0)$, $x_j(t_e)$, and τ_j.

Another result that will be of much use to us is an expression giving the time interval between any two points a and b on an exponential curve. This can be easily derived from Eq. (14-40).

$$t_{ab} \equiv t_b - t_a = \tau_j \ln \frac{x(t_a) - x_j(t_e)}{x(t_b) - x_j(t_e)} \tag{14-44}$$

This equation is valid regardless of the sign of τ_j. However, if $\tau_j < 0$, it is more convenient to rewrite Eq. (14-44) in the form

$$t_{ab} \equiv t_b - t_a = -\tau_j \ln \frac{x(t_b) - x_j(t_e)}{x(t_a) - x_j(t_e)} \tag{14-45}$$

Finally, it is sometimes convenient to be able to determine by inspection the slope at a given point on an exponential. This is given earlier by Eq. (14-35), which we rewrite into the following form:

$$x'(t) = \frac{x_j(t_e) - x(t)}{\tau_j} \tag{14-46}$$

If we let $t = t_0$ in Eq. (14-46) and combine it with Eq. (14-40), we obtain

$$x(t) = x_j(t_e) - \tau_j x'(t_0)e^{-[(t-t_0)/\tau_j]} \tag{14-47}$$

Equation (14-47) represents an alternate way of determining the equation of an exponential in terms of the time constant τ_j, the equilibrium state $x_j(t_e)$, and the slope $x'(t_0)$ at the initial time t_0. This shows that it is always possible to draw an exponential with any prescribed initial slope, any equilibrium state, and any time constant.

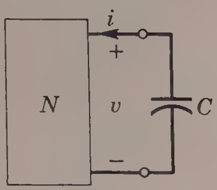
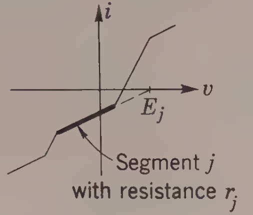
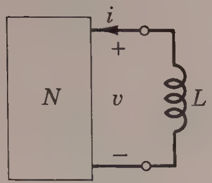
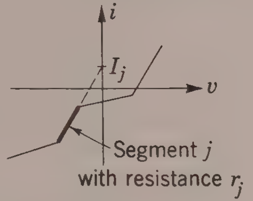
For future reference the various properties characterizing an exponential waveform and their relationships to the solution of first-order networks are summarized in Table 14-1. With the help of this table, we can obtain the complete solution of any piecewise-linear first-order network by inspection. The basic procedure can be outlined as follows:

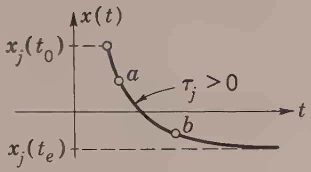
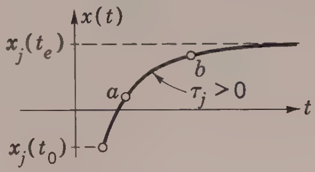
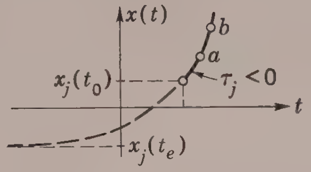
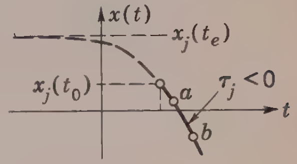
1. Label the segments of the DP plot from 1 to n, starting from the leftmost segment.
2. Locate the pertinent segment j containing the initial state $x_j(t_0)$.†
3. Using the relationships given in Table 14-1, determine the equilibrium state $x_j(t_e)$ and the time constant τ_j corresponding to segment j of the DP plot.
4. With $x_j(t_0)$, $x_j(t_e)$, and τ_j known, sketch the exponential waveform and specify the corresponding equation $x(t)$ for $t \geq 0$.
5. If the value of $x(t)$ obtained in step 4 stays within the interval of definition of segment j for all times $t \geq t_0$, then the complete solution consists of this single exponential time function.
6. If the value of $x(t)$ exceeds the interval of definition of segment j, then we must determine the time $t = t_1$ where $x(t)$ reaches the terminal state (breakpoint) of segment j. The exponential solution is then valid only for $t_0 \leq t \leq t_1$.
7. To find the continuation of the solution for $t \geq t_1$, determine the transition state with the help of the inertia postulate and the principles of conservation of charge and flux linkages.
8. With the transition state obtained in step 7 as the new initial state, repeat steps 2 to 7 as many times as necessary until the solution $x(t)$ is found for all times $t \geq t_0$. The solution asso-

---

† The value $x_j(t_0)$ represents either the initial voltage v on the DP plot for the capacitor or the initial current i on the DP plot for the inductor. Observe further that whereas $v(t_0) = v_C(t_0)$, $i(t_0) = -i_L(t_0)$.

TABLE 14-1 *General properties characterizing the exponential waveform solutions of first-order networks.*

Network	DP plot	Initial state $x_j(t_0)$	Equilibrium state $x_j(t_e)$	Time constant τ_j	Initial slope $x'(t_0)$
		$v_j(t_0)$ = value of v at initial time $t = t_0$	$v_j(t_e)$ = voltage intercept E_j of segment j (or its extension) of the DP plot	$\tau_j = r_jC$	$v'(t_0) = \dfrac{v_j(t_e) - v(t_0)}{\tau_j}$
		$i_j(t_0)$ = value of i at initial time $t = t_0$	$i_j(t_e)$ = current intercept I_j of segment j (or its extension) of the DP plot	$\tau_j = \dfrac{L}{r_j}$	$i'(t_0) = \dfrac{i_j(t_e) - i(t_0)}{\tau_j}$

	Positive time constant		Negative time constant	
	Case 1. $x_j(t_0) > x_j(t_e)$	Case 2. $x_j(t_0) < x_j(t_e)$	Case 1. $x_j(t_0) > x_j(t_e)$	Case 2. $x_j(t_0) < x_j(t_e)$
General form of exponential waveform				
Time interval $t_{ab} \equiv t_b - t_a$ between points a and b	$t_{ab} = \tau_j \ln \dfrac{x(t_a) - x_j(t_e)}{x(t_b) - x_j(t_e)}$		$t_{ab} = -\tau_j \ln \dfrac{x(t_b) - x_j(t_e)}{x(t_a) - x_j(t_e)}$	
General equation	In terms of $x_j(t_0)$, $x_j(t_e)$, and τ_j: $x(t) = x_j(t_e) + [x_j(t_0) - x_j(t_e)]e^{-[(t-t_0)/\tau_j]}$ In terms of $x'(t_0)$, $x_j(t_e)$, and τ_j: $x(t) = x_j(t_e) - \tau_j x'(t_0)e^{-[(t-t_0)/\tau_j]}$			

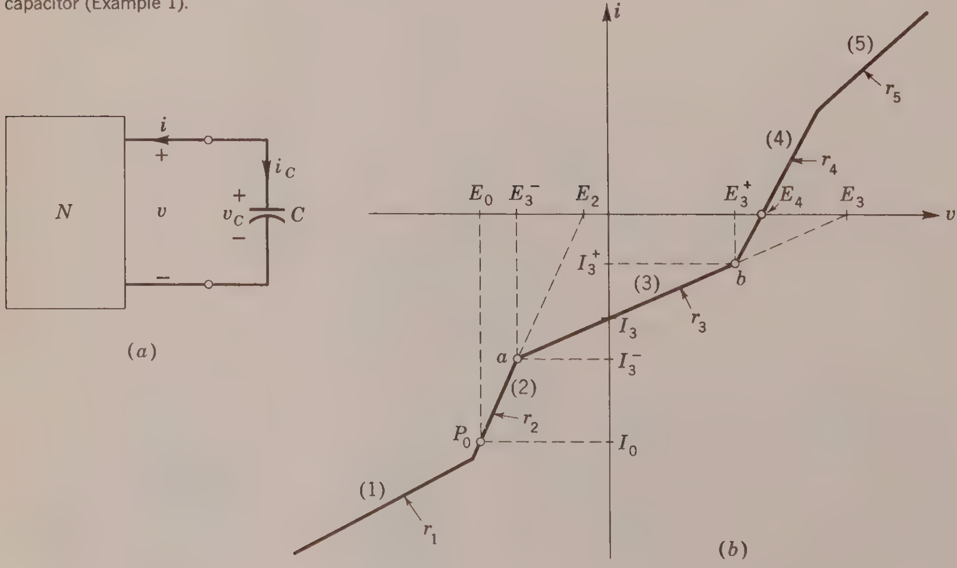
ciated with any piecewise-linear segment of the DP plot will always be an exponential waveform, provided that each segment of the DP plot involved has a finite and nonzero slope.

EXAMPLES

1. Consider the network shown in Fig. 14-16*a* where the DP plot of the resistive subnetwork N is shown in Fig. 14-16*b*. If the initial voltage across the capacitor at $t = t_0$ is $v_C(t_0) = E_0$ (the value E_0 is labeled on the v axis), find the capacitor voltage $v_C(t)$ and the capacitor current $i_C(t)$ for all time $t \geq t_0$.

   Since $v = v_C$ we have $v(t_0) = E_0$, and hence N is operating initially on P_0 of segment 2. This means that at $t = t_0$ (and at all times $t \geq t_0$ where the operating point remains on segment 2), the DP plot of N appears as a straight line to the capacitor as shown in Fig. 14-17*a*. This follows because until such time as the operating point reaches breakpoint a of segment 2, the capacitor does not know that the DP plot of N is nonlinear. What the capacitor "experienced" here is analogous to what a man experiences driving on a straight but signless road on a foggy day. If the road eventually makes an abrupt change in direction, the man will not know it until he arrives at the turnoff point. However, even though he did not have this information, he can correctly predict his "present" position by

Fig. 14-16. A first-order network with a monotonically increasing DP plot and a linear capacitor (Example 1).



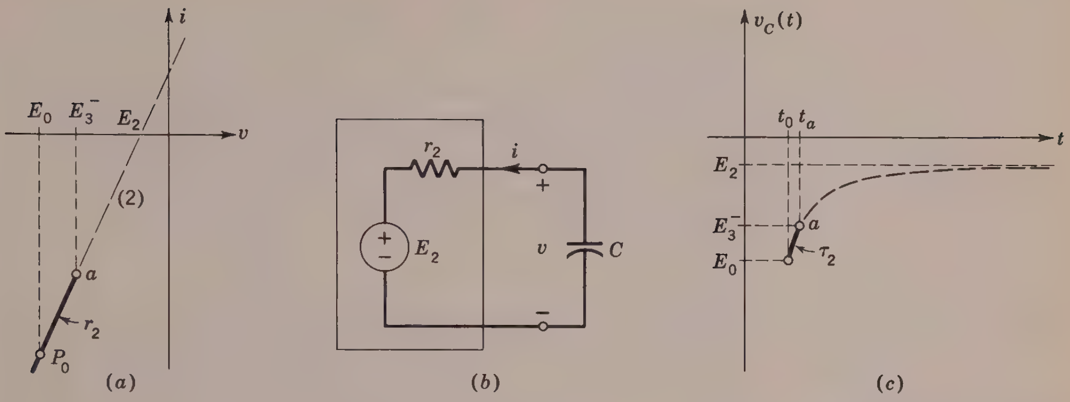


Fig. 14-17. The iterative linear network for determining $v_C(t)$ during the time interval where N is operating on segment 2.

assuming that the road is perfectly straight up to infinity, provided he knows his initial position and his speed. Likewise, we can determine $v_C(t)$ for all times where the operating point remains on segment 2 by replacing N with its iterative linear network as shown in Fig. 14-17*b*.

The solution can now be obtained by determining the initial state, the equilibrium state, and the time constant corresponding to segment 2.

The initial state is given by $v(t_0) = E_0$. The equilibrium state can be determined by inspection of either Fig. 14-17*a* or *b*; namely, $v_2(t_e) = E_2$.† From Table 14-1 we obtain the time constant $\tau_2 = r_2C$ and the associated solution

$$v(t) = E_2 + (E_0 - E_2)e^{-(t-t_0)/r_2C} \tag{14-48}$$

for $t \geq t_0$. The solution can also be sketched directly as shown in Fig. 14-17*c*. This is done by plotting an exponential waveform with the time constant τ_2 starting at $v = E_0$ and approaching an asymptote at $v = E_2$.

An examination of the DP plot of Fig. 14-16*b* shows that segment 2 is defined only up to $v = E_3^-$. Therefore, only the solid portion of the exponential waveform shown in Fig. 14-17*c* constitutes the valid solution. To determine the time t_a when the solution arrives at breakpoint a, we obtain from Eq. (14-44)

$$t_a = t_0 + (r_2C) \ln \frac{E_0 - E_2}{E_3^- - E_2} \tag{14-49}$$

Since point a represents a terminal state, the capacitor must make a joint decision with the network N as to where to go next so that both C and N will remain "happy." In other words, we must find the transition state. As far as the capacitor

---

† Recall that the equilibrium state for the capacitor case is equal to the voltage intercept of the pertinent segment or its extension. It is also equal to the voltage across the capacitor terminals when the capacitor is replaced by an open circuit. For identification purposes, a subscript corresponding to the pertinent segment will be associated with the variable v.

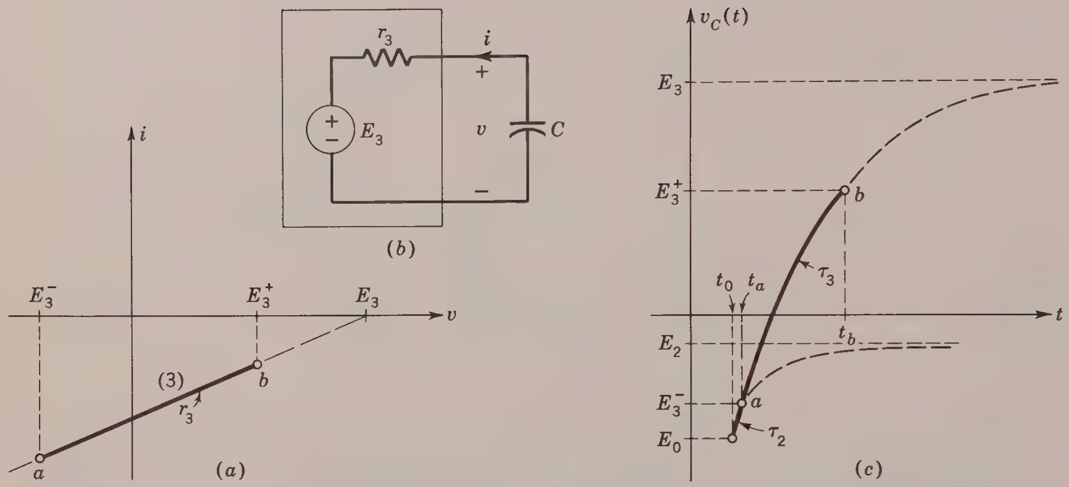


Fig. 14-18. The iterative linear network for determining $v_C(t)$ during the time interval when N is operating on segment 3.

is concerned, we have $i = -C\,dv/dt$ or $dv/dt = -i/C$. Since point a is in the lower half-plane, the value of i at this point is negative; therefore, $dv/dt > 0$ at the terminal state a. This means that to make the capacitor "happy," its voltage $v(t)$ at $t = t_a$ must continue to increase. But as far as N is concerned, the network must at all times be operating at a point on the DP plot. It is possible to satisfy both C and N by continuing to segment 3. Point a of segment 3 is, therefore, the transition state.

At $t = t_a^+$, and at all times thereafter while the operating point remains on segment 3, the capacitor sees the linear DP plot shown in Fig. 14-18*a*. Therefore, we can replace N by its iterative Thévenin equivalent circuit as shown in Fig. 14-18*b*. The initial state now is given by $v_3(t_a) = E_3^-$. The equilibrium state is given by $v_3(t_e) = E_3$, and the time constant is simply $\tau_3 = r_3C$. The corresponding solution for $t \geq t_a$ is given by

$$v(t) = E_3 + (E_3^- - E_3)e^{-(t-t_a)/r_3C} \tag{14-50}$$

and the waveform is readily sketched as shown in Fig. 14-18*c*.

However, upon arriving at breakpoint b of the DP plot of Fig. 14-16*b*, the terminal state of segment 3 is reached, and the solution given by Eq. (14-50) is, therefore, valid only in the time interval $t_a \leq t < t_b$, where

$$t_b = t_a + (r_3C)\ln\frac{E_3^- - E_3}{E_3^+ - E_3} \tag{14-51}$$

To determine the transition state again, we find that $dv/dt = -i/C$ is positive at b because $i < 0$ at this point. We conclude by the same reasoning as before that the operating point will continue to segment 4.

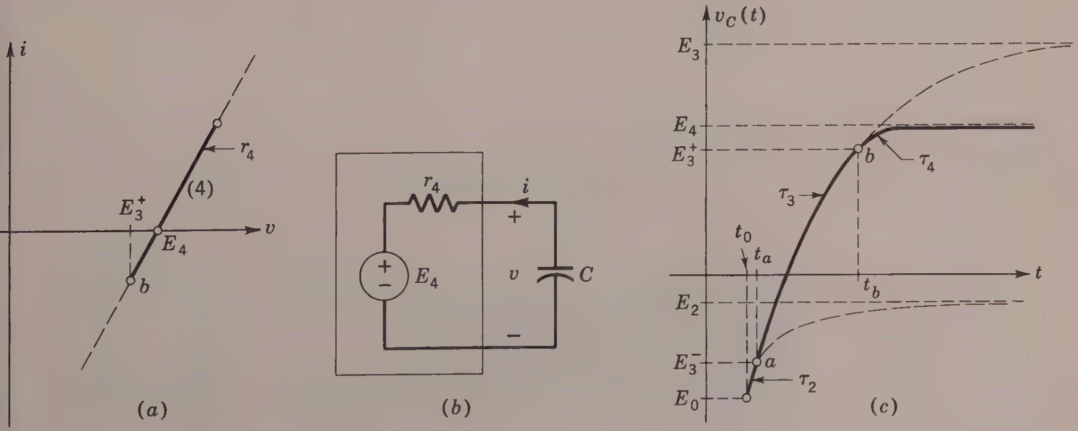
On segment 4, the capacitor sees the linear DP plot shown in Fig. 14-19*a* and the iterative Thévenin network shown in Fig. 14-19*b*. The initial state in this case is given by $v(t_b) = E_3^+$, the equilibrium state is given by $v_4(t_e) = E_4$, and the time constant τ_4 is equal to r_4C. The solution for $t \geq t_b$ is, therefore, given by

$$v(t) = E_4 + (E_3^+ - E_4)e^{-(t-t_b)/r_4C} \tag{14-52}$$

and the corresponding waveform is sketched as shown in Fig. 14-19*c*. Since it will take an infinite time interval for this exponential to reach the equilibrium value E_4, and since $v = E_4$ is contained within segment 4, the solution given by Eq. (14-52) is valid for $t_b \leq t < \infty$. Since $v_C = v$, the complete solution is given by

$$v_C(t) = \begin{cases} E_2 + (E_0 - E_2)e^{-(t-t_0)/r_2C} & t_0 \leq t < t_a \\ E_3 + (E_3^- - E_3)e^{-(t-t_a)/r_3C} & t_a \leq t < t_b \\ E_4 + (E_3^+ - E_4)e^{-(t-t_b)/r_4C} & t_b \leq t < \infty \end{cases} \tag{14-53}$$

The corresponding waveform for $v_C(t)$ is identical with that shown in Fig. 14-19*c*, and is replotted in Fig. 14-20*a* for future reference. Observe that the two adjacent exponential waveforms at each transition time t_a or t_b have identical slopes. This observation can be verified by evaluating $v_C'(t)$ from Eq. (14-53) at t_a and t_b. The fact that the slopes are identical at t_a or t_b can also be proved from the relationship $i = C\,dv_C/dt$. Since C is a constant, the current i will exhibit a jump if dv_C/dt at t_a^- and t_a^+ are not identical. But we have just found that the operating



(*a*) (*b*) (*c*)

Fig. 14-19. The iterative linear network for determining $v_C(t)$ during the time interval where N is operating on segment 4.

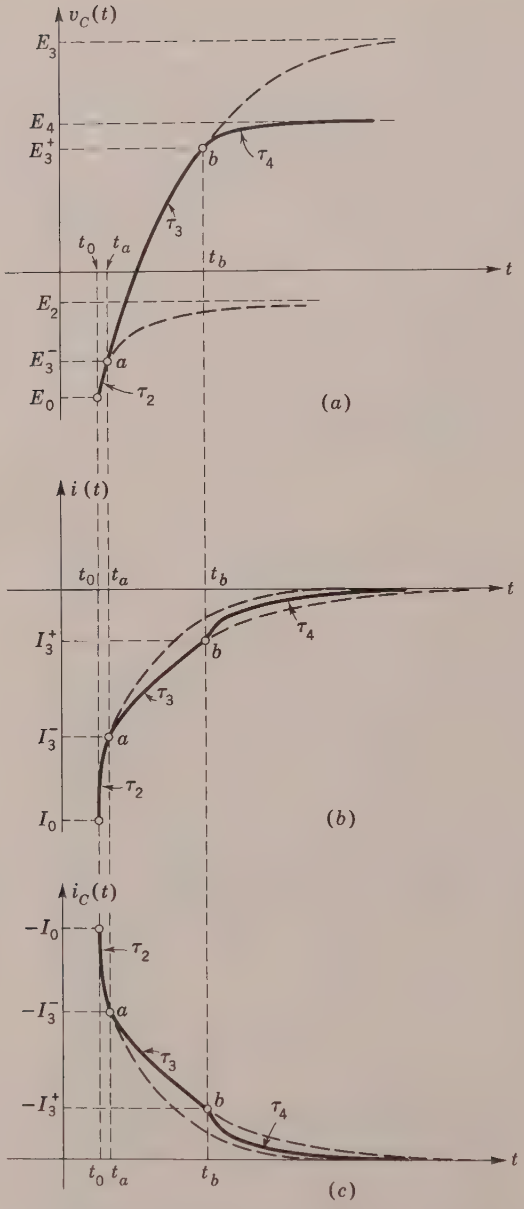


Fig. 14-20. The complete solution waveforms $v_C(t)$, $i(t)$, and $i_C(t)$ for Example 1.

point traverses the DP plot of Fig. 14-16*b* *continuously* from point P_0 to the equilibrium state at $v = E_4$, and therefore cannot exhibit any instantaneous jump.

To complete the solution of this problem, let us determine the capacitor current $i_C(t)$. Since $i_C = -i$, it suffices for us to determine $i(t)$. There are two methods for determining $i(t)$. The first method consists of evaluating

$$i(t) = -C\frac{dv_C}{dt} \tag{14-54}$$

directly from $v_C(t)$ given in Eq. (14-53). The result is

$$i(t) = \begin{cases} \dfrac{1}{r_2}(E_0 - E_2)e^{-(t-t_0)/r_2C} & t_0 \le t < t_a \\ \dfrac{1}{r_3}(E_3^- - E_3)e^{-(t-t_a)/r_3C} & t_a \le t < t_b \\ \dfrac{1}{r_4}(E_3^+ - E_4)e^{-(t-t_b)/r_4C} & t_b \le t < \infty \end{cases} \tag{14-55}$$

This solution could have been obtained directly from $v_C(t)$, as shown in Fig. 14-20*a*, by graphical differentiation. In either case, the resulting waveform for $i(t)$ is shown in Fig. 14-20*b*, and the desired solution $i_C(t)$ is simply the negative of $i(t)$ as shown in Fig. 14-20*c*. The above method for determining $i_C(t)$ is conceptually rather straightforward. However, it is a circuitous method because it is necessary to determine $v_C(t)$ first. In problems where only $i_C(t)$ is sought, it will be much easier to use an alternate method described as follows.

The alternate method for determining the capacitor current $i_C(t)$ by inspection is based upon the following two important properties of exponential solutions of iterative linear networks:

Property 1: The derivative of any exponential time function is also an exponential with an identical time constant.

Property 2: The capacitor current $i_C(t)$ is always equal to zero when it is in equilibrium. Therefore, $i_j(t_e) = 0$ on any segment.

In view of these properties we may conclude that since $v_C(t)$ always consists of exponential segments, so does $i(t)$. Moreover, the transition times for $v_C(t)$ and $i(t)$ will always be identical because $v_C(t)$ and $i(t)$ have identical time constants. This means that we can always determine the initial state $i(t_0)$, the equilibrium state $i_j(t_e)$ (which is always zero because of property 2), and the time constant $\tau_j = r_jC$. Therefore, once the sequence of operating segments, henceforth called the *dynamic route,* is known, the corresponding solution $i(t)$ can be determined by inspection. In this example, the dynamic route starting at P_0 and

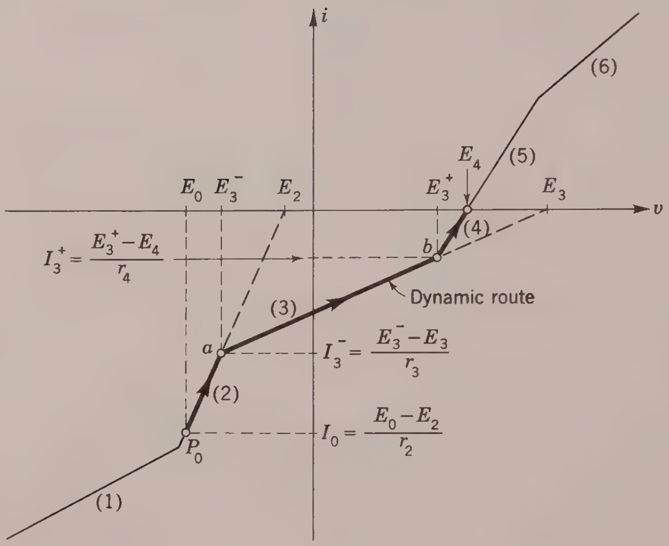

i
(6)
E_4
(5)
E_0
E_3^-
E_2
E_3^+
E_3
v
(4)
b
$I_3^+ = \frac{E_3^+ - E_4}{r_4}$
Dynamic route
(3)
a
$I_3^- = \frac{E_3^- - E_3}{r_3}$
(2)
$I_0 = \frac{E_0 - E_2}{r_2}$
P_0
(1)

Fig. 14-21. The solution $i(t)$ can be determined by inspection of the dynamic route along the DP plot.

terminating at the equilibrium state $v = E_4$ or $i = 0$ is easily determined as before and is shown by the bold lines in Fig. 14-21.

To determine $i(t)$, we start with segment 2, and by inspection of the dynamic route obtain $i(t_0) = I_0$, $i_2(t_e) = 0$, and $\tau_2 = r_2C$. The solution is, therefore, given by

$$i(t) = I_0 e^{-(t-t_0)/r_2C} \qquad t_0 \leq t < t_a \tag{14-56}$$

where

$$t_a = t_0 + (r_2C) \ln \frac{I_0 - 0}{I_3^- - 0} \tag{14-57}$$

Similarly, on segment 3 we obtain $i(t_a) = I_3^-$, $i_3(t_e) = 0$, and $\tau_3 = r_3C$. The corresponding solution is

$$i(t) = I_3^- e^{-(t-t_a)/r_3C} \qquad t_a \leq t < t_b \tag{14-58}$$

where

$$t_b = t_a + (r_3C) \ln \frac{I_3^- - 0}{I_3^+ - 0} \tag{14-59}$$

Finally, on segment 4 we obtain $i(t_b) = I_3^+$, $i_4(t_e) = 0$, and $\tau_4 = r_4C$. The corresponding solution is

$$i(t) = I_3^+ e^{-(t-t_b)/r_4C} \qquad t_b \leq t < \infty \tag{14-60}$$

Applying simple analytic geometry to the DP plot shown in Fig. 14-21, we easily see that

$$I_0 = \frac{(E_0 - E_2)}{r_2} \tag{14-61}$$

$$I_3^- = \frac{(E_3^- - E_3)}{r_3} = \frac{(E_3^- - E_2)}{r_2} \tag{14-62}$$

$$I_3^+ = \frac{(E_3^+ - E_4)}{r_4} = \frac{(E_3^+ - E_3)}{r_3} \tag{14-63}$$

If we substitute I_0, I_3^-, and I_3^+ in Eqs. (14-56) to (14-60), we obtain Eq. (14-55), thus verifying the validity of the above method.

Comparing the above methods, we see clearly that the alternate method is computationally much simpler than the earlier method.

2. Find the inductor current $i_L(t)$ and the inductor voltage $v_L(t)$ of the network shown in Fig. 14-22*a*, corresponding to an initial current $i_L(t_0) = -I_0$. The DP plot of the resistive subnetwork N is current-controlled and is shown in Fig. 14-22*b*.

   Since the DP plot is given in the v-i plane, it is more convenient for us to choose i instead of i_L as the state variable. Since $i(t_0) = I_0$, the network is initially operating on segment 4 as shown by the point P_0 in Fig. 14-22*b*. As before, it is neces-

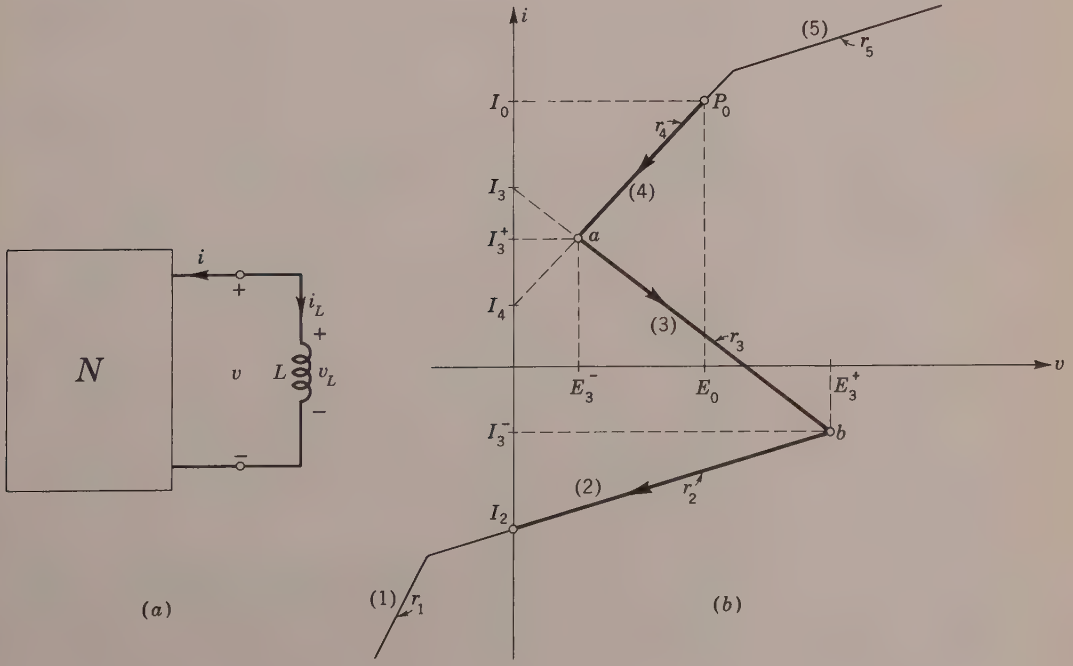
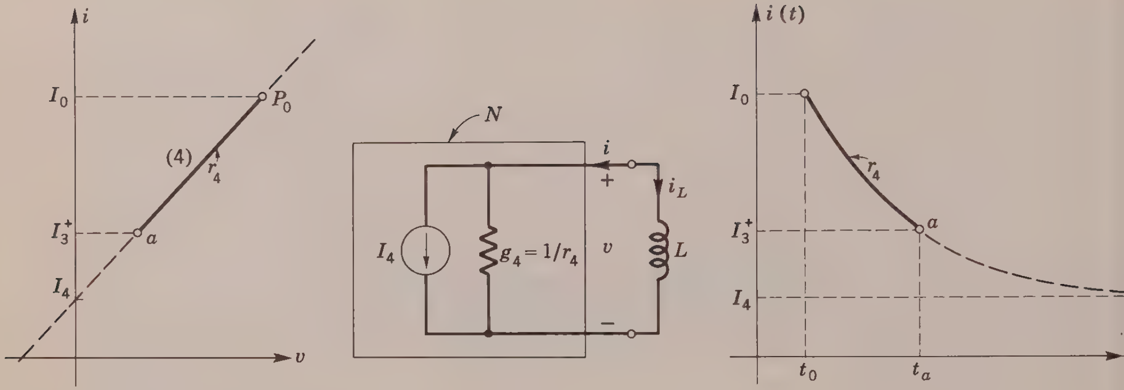


Fig. 14-22. A first-order network with a current-controlled DP plot and a linear inductor (Example 2).



(a) (b) (c)

Fig. 14-23. The iterative linear network for determining $i(t)$ during the time interval where N is operating on segment 4.

sary to determine the dynamic route along the DP plot. This can be easily found from the basic relationship $di/dt = -v/L$. Observe that $di/dt < 0$ whenever $v > 0$, and vice versa. This means that the current i in the right half-plane must be decreasing with time, and the current i in the left half-plane must be increasing with time. Since the initial segment is in the right half-plane, the dynamic route will move down segment 4, as shown in Fig. 14-22*b*, until it arrives at the terminal state *a*, which is still in the right half-plane. Therefore, the dynamic route must continue to decrease along segment 3. It will cross the v axis and eventually arrive at the terminal state *b*. Since point *b* is still in the right half-plane, the dynamic route will continue to decrease along segment 2 until it arrives at the equilibrium state $i = I_2$, as shown in Fig. 14-22*b*.

By arguments dual to those used in the preceding example, we observe that during the entire time interval while the network is operating in segment 4, the inductor sees a straight DP plot as shown in Fig. 14-23*a*. The resistive subnetwork N can, therefore, be replaced by its iterative linear equivalent circuit as shown in Fig. 14-23*b*. The initial state is given by $i(t_0) = I_0$, and the equilibrium state is given by $i_4(t_e) = I_4$. Since the energy-storage element is an inductor, the time constant is given by $\tau_4 = L/r_4$. The solution corresponding to segment 4 is, therefore, given by

$$i(t) = I_4 + (I_0 - I_4)e^{-r_4(t-t_0)/L} \tag{14-64}$$

where $t \geq t_0$. Alternately, we can obtain the solution graphically by inspection as shown in Fig. 14-23*c*. Obviously, this solution is valid only up to the time t_a, where

$$t_a = t_0 + \frac{L}{r_4} \ln \frac{I_0 - I_4}{I_3^+ - I_4} \tag{14-65}$$

To find $i(t)$ for $t \geq t_a$, we observe that the DP plot can be replaced by the straight line shown in Fig. 14-24*a* and the corresponding iterative linear network as shown in Fig. 14-24*b*. The initial and the equilibrium states are given, respectively, by $i(t_a) = I_3^+$ and $i_3(t_e) = I_3$. Since the slope of this line is negative, we obtain a negative time constant $\tau_3 = -L/|r_3|$. The corresponding solution is given by

$$i(t) = I_3 + (I_3^+ - I_3)e^{|r_3|(t-t_a)/L} \tag{14-66}$$

where $t \geq t_a$. Alternately, this solution can be obtained graphically as shown in Fig. 14-24*c*. Observe that we have used the properties that for a negative time constant the exponential through the initial state $i = I_3^+$ is asymptotic to the horizontal line through the equilibrium state $i = I_3$, and that if we move back one time constant to $t = t_a - |\tau_3|$, the exponential waveform will reach 37 percent of $(I_3 - I_3^+)$ from its equilibrium state at $i = I_3$. Another important clue for drawing this exponential is to notice that the slopes of the preceding and the present exponential waveforms at $t = t_a$ are continuous. This follows because the dynamic route does not show a jump in the inductor voltage $v_L = -L\, di/dt$ at $t = t_a$. The above solution is valid only over the time interval (t_a, t_b) where

$$t_b = t_a + \frac{L}{|r_3|} \ln \frac{I_3^- - I_3}{I_3^+ - I_3} \tag{14-67}$$

For $t \geq t_b$, the inductor sees the straight DP plot shown in Fig. 14-25*a*, and the corresponding iterative linear network is

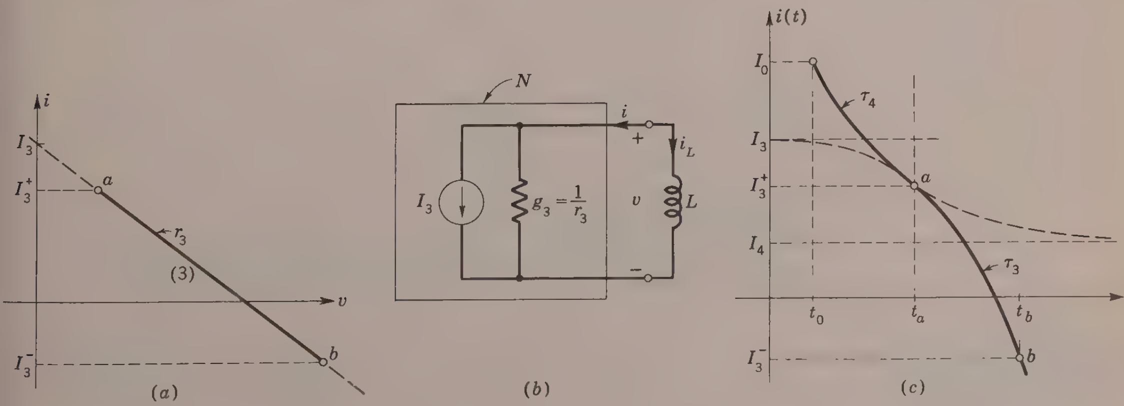


Fig. 14-24. The iterative linear network for determining $i(t)$ during the time interval where N is operating on segment 3.

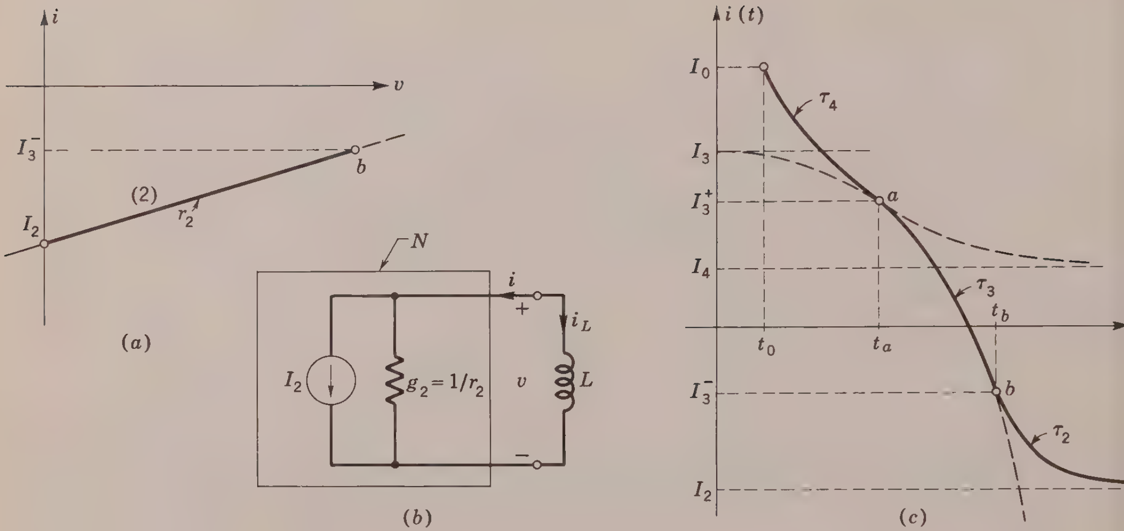


Fig. 14-25. The iterative linear network for determining $i(t)$ during the time interval where N is operating on segment 2.

shown in Fig. 14-25*b*. The initial and equilibrium states are given, respectively, by $i(t_b) = I_3^-$ and $i_2(t_e) = I_2$. The time constant is once again positive and is given by $\tau_2 = L/r_2$. From this information we obtain the solution

$$i(t) = I_2 + (I_3^- - I_2)e^{-r_2(t-t_b)/L} \tag{14-68}$$

where $t \geq t_b$. This solution can also be obtained graphically as shown in Fig. 14-25*c*. Observe again that the slope at $t = t_b$ must be continuous. Otherwise, the inductor voltage will jump here, thus contradicting the continuous dynamic route shown in Fig. 14-22*b*.[1]

Since $i_L = -i$, the complete solution $i_L(t)$ is obtained from Eqs. (14-64) to (14-68) as follows:

$$i_L(t) = \begin{cases} -I_4 - (I_0 - I_4)e^{-r_4(t-t_0)/L} & t_0 \leq t < t_a \\ -I_3 - (I_3^+ - I_3^-)e^{|r_3|(t-t_a)/L} & t_a \leq t < t_b \\ -I_2 - (I_3^- - I_2)e^{-r_2(t-t_b)/L} & t_b \leq t < \infty \end{cases} \tag{14-69}$$

This is plotted in Fig. 14-26*a*.

It is now a simple matter to find $v_L(t)$. We can use the relationship $v_L = L\, di_L/dt$ and Eq. (14-69) to obtain

$$v_L(t) = \begin{cases} r_4(I_0 - I_4)e^{-r_4(t-t_0)/L} & t_0 \leq t < t_a \\ -|r_3|(I_3^+ - I_3^-)e^{|r_3|(t-t_a)/L} & t_a \leq t < t_b \\ r_2(I_3^- - I_2)e^{-r_2(t-t_b)/L} & t_b \leq t < \infty \end{cases} \tag{14-70}$$

---

[1] We shall say a dynamic route is continuous if it does not exhibit any instantaneous jumps.

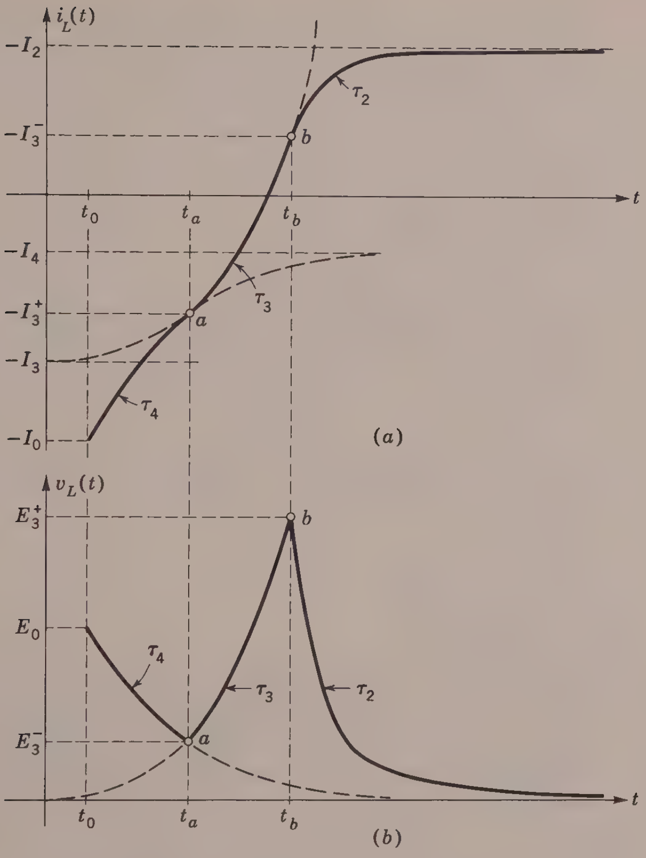


Fig. 14-26. The complete solution waveforms $i_L(t)$ and $v_L(t)$ for Example 2.

We can also obtain this solution by a graphical differentiation procedure as shown in Fig. 14-26*b*. In either case these methods require the determination of $i_L(t)$ first.

It is possible to avoid this circuitous approach by taking advantage of the property that $v_L(t)$ must consist also of exponential waveforms having time constants identical with $i(t)$ over the same time intervals. Therefore, all we need to know are the initial state, the equilibrium state, and the time constant over each segment of the dynamic route. For example, corresponding to segment 4 of Fig. 14-22*b*, we have $v(t_0) = E_0$, $v_4(t_e) = 0$,† and $\tau_4 = L/r_4$. Therefore, we have

$$v(t) = E_0 e^{-r_4(t-t_0)/L} \qquad t_0 \leq t < t_a \tag{14-71}$$

where

$$t_a = t_0 + \frac{L}{r_4} \ln \frac{E_0}{E_3^-} \tag{14-72}$$

---

† Observe that the voltage across an inductor in equilibrium is always equal to zero.

Corresponding to segment 3 of the dynamic route, we obtain the initial state $v(t_a) = E_3^-$, the equilibrium state $v_3(t_e) = 0$, and $\tau_3 = -L/|r_3|$. Therefore, we have

$$v(t) = E_3^- e^{|r_3|(t-t_a)/L} \qquad t_a \leq t < t_b \tag{14-73}$$

where

$$t_b = t_a + \frac{L}{|r_3|} \ln \frac{E_3^+}{E_3^-} \tag{14-74}$$

Finally, corresponding to segment 2 of the dynamic route, the initial state $v(t_b) = E_3^+$, the equilibrium state $v_2(t_e) = 0$, and the time constant $\tau_2 = L/r_2$ are obtained. Therefore, we have

$$v(t) = E_3^+ e^{-r_2(t-t_b)/L} \qquad t_b \leq t < \infty \tag{14-75}$$

By simple analytic geometry it is easy to derive from Fig. 14-22*b* the following identities:

$$E_0 = r_4(I_0 - I_4) \tag{14-76}$$

$$E_3^- = -|r_3|(I_3^+ - I_3) \tag{14-77}$$

$$E_3^+ = r_2(I_3^- - I_2) \tag{14-78}$$

Substituting these identities into Eqs. (14-71) to (14-75), we obtain the same result as Eq. (14-70). Observe, however, that the present method is computationally much simpler because the equilibrium voltage $v_j(t_e)$ is always zero,[1] and no differentiation is needed.

3. Consider next the network shown in Fig. 14-27*a* with the DP plot shown in Fig. 14-27*b*. If the initial capacitor voltage is given to be $v_C(t_0) = E_0$, find $v_C(t)$ and $i_C(t)$ for all times $t \geq t_0$.

   As usual, the first step is to determine the dynamic route. Since the energy-storage element is a capacitor, we have $dv/dt = -i/C$ and hence, $dv/dt < 0$ whenever $i > 0$ and $dv/dt > 0$ whenever $i < 0$. From this, we can formulate the following general rule: *the capacitor voltage must always decrease along the DP plot in the upper half-plane and increase along the DP plot in the lower half-plane.* Therefore, starting with the initial state P_0 on segment 2, the dynamic route must move up along this segment as shown in Fig. 14-27*b*. When it reaches the terminal state *a*, the capacitor voltage must continue to increase because point *a* is still in the lower half-plane. But it cannot continue to segment 3 because the voltage is decreasing there. The only way out of this dilemma is for the dynamic route to

---

[1] This is dual to the capacitor case where the equilibrium current is always zero.

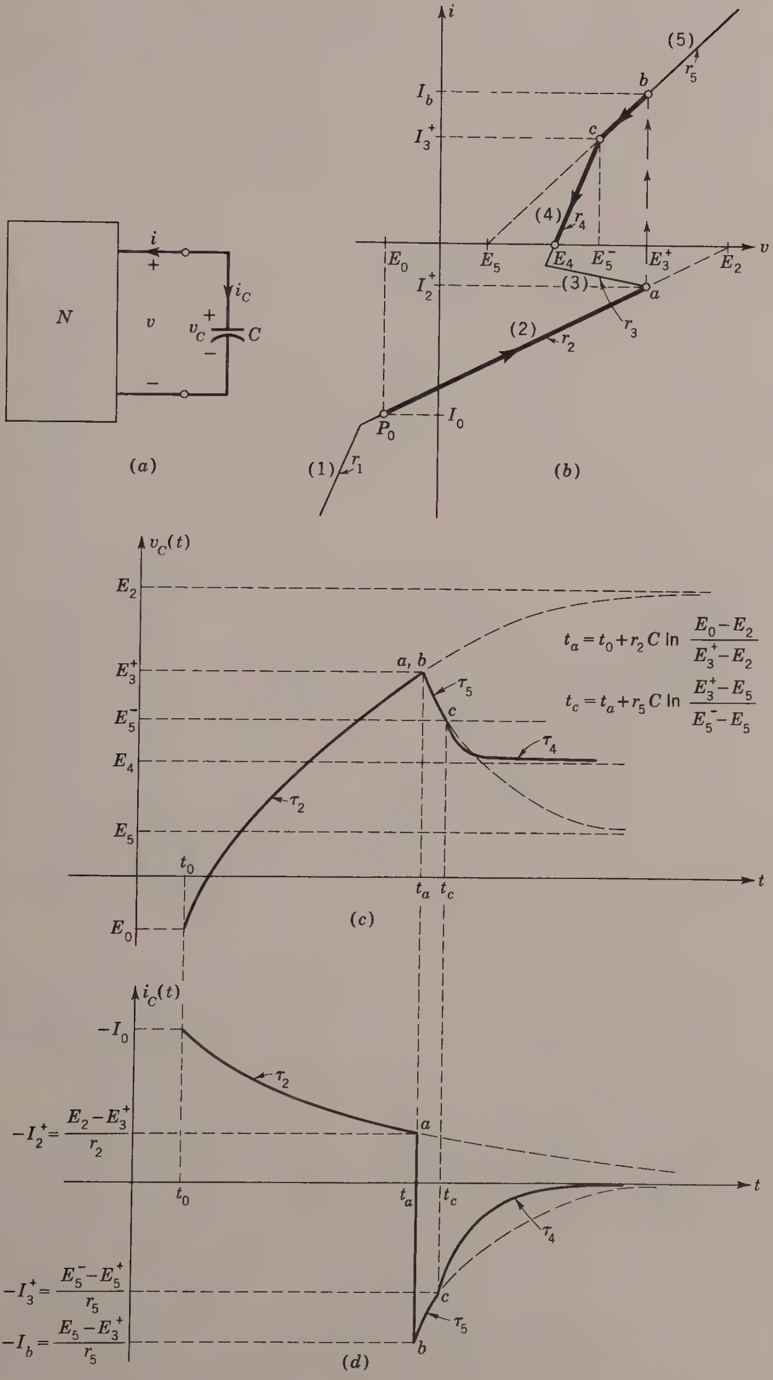


Fig. 14-27. A first-order network with a current-controlled DP plot and its associated solution waveforms (Example 3).

jump up instantaneously along the vertical line shown in Fig. 14-27*b*.[1] Having established the transition state at point *b* (which is now in the upper half-plane), the capacitor voltage must begin to decrease along segment 5 and continue down along segment 4 until it arrives at the equilibrium state $v = E_4$. The complete dynamic route is shown by the solid paths in Fig. 14-27*b*. The next step is to determine the exponential solution associated with each segment.

A review of the detailed procedure in the preceding example shows that a number of steps can easily be done mentally and can therefore be omitted. For example, the iterative linear networks used in the preceding example are really unnecessary. They were used earlier in order to illustrate the main concept in a clearer perspective. The complete solution can be determined by inspection of the dynamic route shown in Fig. 14-27*b*. For convenience the pertinent information for each segment is summarized in Table 14-2. The complete solutions $v_C(t)$ and $i_C(t)$ for all times $t \geq t_0$ are easily obtained from the data in this table and are shown, respectively, in Fig. 14-27*c* and *d*. Observe that, unlike the previous examples, the slope of $v_C(t)$ at $t = t_a$ is discontinuous.[2] Since $i_C(t) = C\ dv_C/dt$, the capacitor current $i_C(t)$ must exhibit a jump here. This clearly agrees with the dynamic route shown in Fig. 14-27*b*. Observe further that at $t = t_c$, the slope of $v_C(t)$ is continuous, and the corresponding current $i_C(t)$ is, therefore, continuous. This agrees again with the dynamic route. It should now be clear that our method in the present section provides many independent self-checking features. Any inconsistency that may arise means that an error has been committed, and it will be necessary to uncover the error before proceeding further.

4. As a final example consider the network shown in Fig. 14-28*a* with the DP plot shown in Fig. 14-28*b*. The problem is to find $i(t)$ and $v_L(t)$ for all times $t \geq t_0$ corresponding to an initial inductor current $i_L(t_0) = -I_0$.

TABLE 14-2 ***Pertinent information for describing the dynamic route of Example 3.***

Segment	Initial state		Equilibrium state		Time constant
	$v(t_0)$	$i(t_0)$	$v_j(t_e)$	$i_j(t_e)$	τ_j
2	E_0	I_0	E_2	0	r_2C
5	E_3^+	I_b	E_5	0	r_5C
4	E_5^-	I_3^+	E_4	0	r_4C

---

[1] The jump must be vertical because of the principle of the conservation of charge.

[2] Whenever the voltage across a capacitor or the current in an inductor exhibits two distinct slopes at any time $t = t_j$, the corresponding capacitor current or inductor voltage must exhibit a jump and is therefore undefined at t_j. Accordingly, it becomes necessary to split the time t_j into t_j^- and t_j^+, where t_j^- and t_j^+ are infinitesimally less than and greater than t_j, respectively.

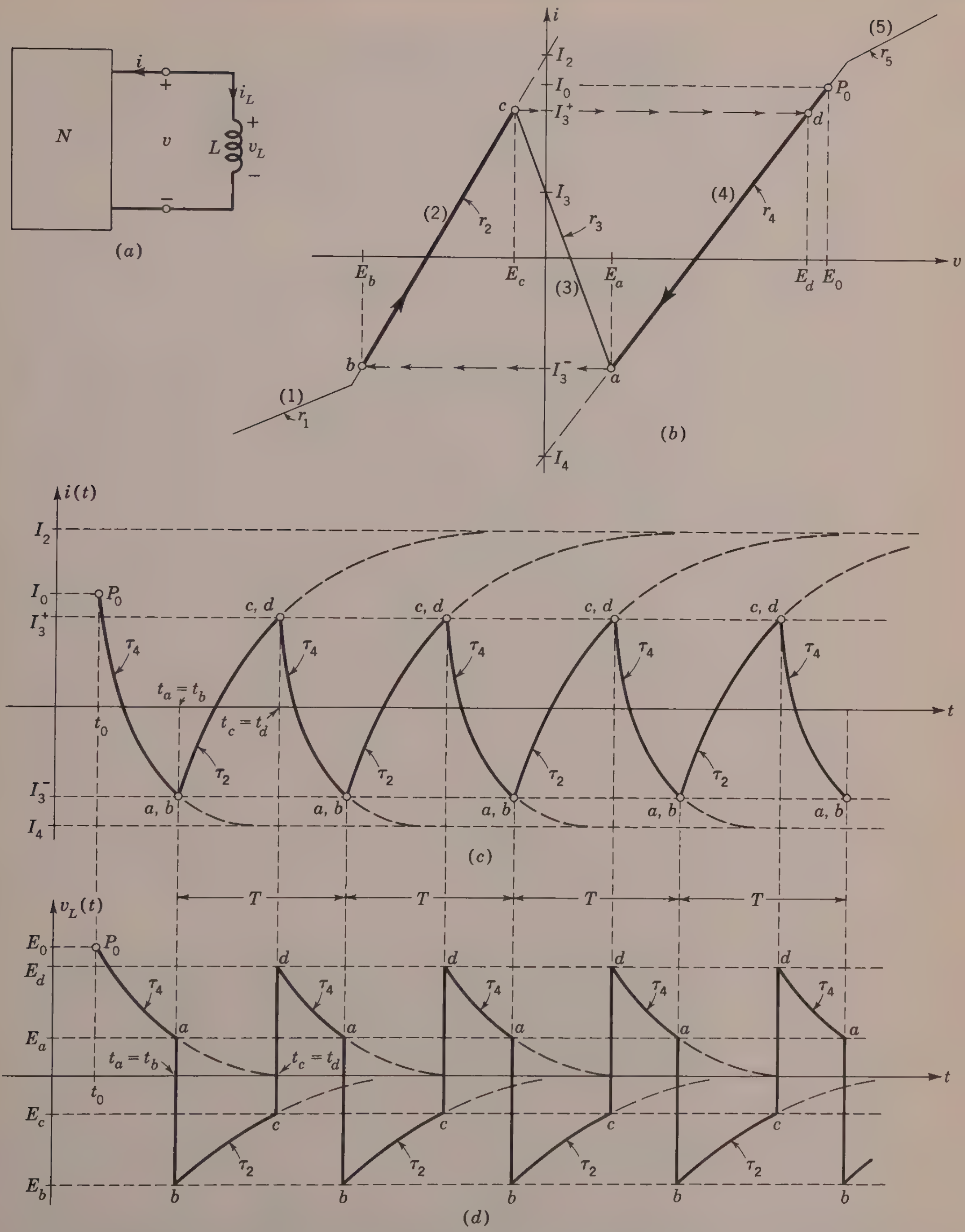


Fig. 14-28. A first-order network with a voltage-controlled DP plot and its associated solution waveforms (Example 4).

TABLE 14-3 *Pertinent information for describing the dynamic route of Example 4.*

Segment	Initial state		Equilibrium state		Time constant
	$i(t_0)$	$v(t_0)$	$i_j(t_e)$	$v_j(t_e)$	τ_j
4	I_0	E_0	I_4	0	$\frac{L}{r_4}$
2	I_3^-	E_b	I_2	0	$\frac{L}{r_2}$

Applying the general rule that the current $i = -i_L$ through the inductor must decrease along the DP plot in the right half-plane and increase along the DP plot in the left half-plane, we immediately obtain the dynamic route as shown by the solid paths in Fig. 14-28*b*. From this we summarize the pertinent data in Table 14-3.

The complete solutions $i(t)$ and $v_L(t)$ are easily determined from the data in Table 14-3 and are shown, respectively, in Fig. 14-28*c* and *d*. Observe that the resulting waveforms for $i(t)$ and $v_L(t)$ are periodic for all times $t \geq t_b$. The period of oscillation T is readily determined as follows:

$$T = (t_c - t_b) + (t_a - t_d) = \frac{L}{r_2} \ln \frac{I_3^- - I_2}{I_3^+ - I_2} + \frac{L}{r_4} \ln \frac{I_3^+ - I_4}{I_3^- - I_4} \tag{14-79}$$

Observe further that $v_L(t)$ exhibits two jumps because $v_L(t) = -L\, di/dt$, and the slope of $i(t)$ exhibits two discontinuities per cycle. This again agrees with the dynamic route shown in Fig. 14-28*b*.

14-8-2 NETWORKS CONTAINING A LINEAR ENERGY-STORAGE ELEMENT ACROSS A NONLINEAR RESISTIVE NETWORK CHARACTERIZED BY PIECEWISE-LINEAR SEGMENTS WITH ARBITRARY SLOPES

So far we have assumed that the DP plot of the resistive subnetwork contains neither horizontal nor vertical segments. The reason for this assumption is that the time constant becomes either infinite or zero when such segments are present, and Eq. (14-35) is no longer applicable. Fortunately, these special cases turn out to be easier to handle because the iterative linear network becomes a current source when the segment is horizontal and a voltage source when the segment is vertical.

Consider first the case where the energy-storage element is a capacitor. To be specific, let us choose a typical DP plot containing both a horizontal and a vertical segment as shown in Fig. 14-29*a*. Suppose it is required to find $v(t)$ and $i(t)$ for all times $t \geq t_0$ corresponding to $v(t_0) = E_0$.

First we must find the dynamic route. Since the capacitor voltage v must decrease along the DP plot in the upper half-plane, the dynamic route starting at P_0 will move *left* along segment 6 until it arrives at the terminal state a, when it must make an *instantaneous* jump along segment 5 to point b. This jump is due to the fact that point a is in the upper half-plane, and hence the voltage v must continue to decrease toward the left. This behavior is obvious from an inspection of the equivalent network shown in Fig. 14-29*b* where a voltage source E_5 is connected across the capacitor at $t = t_a$. Since point b is in the upper half-plane, the dynamic route must continue to move left, first on segment 4, then to the horizontal segment 3, and finally to segment 2 until it arrives at the equilibrium state $v = E_2$.

We can apply our previous methods to obtain the solutions corresponding to each segment of the DP plot. The only difference here lies in segments 5 and 3. Since the time it takes to move from a to b of the vertical segment 5 is zero, only segment 3 needs further consideration. The equivalent network for this horizontal segment is simply the current source shown in Fig. 14-29*c*. To determine the waveform of $v(t)$ on this segment, we observe that

$$v(t) = \frac{-1}{C}\int_{-\infty}^{t} i(\tau)\,d\tau = \frac{-1}{C}\int_{-\infty}^{t_c} i(\tau)\,d\tau - \frac{1}{C}\int_{t_c}^{t} i(\tau)\,d\tau \qquad (14\text{-}80)$$

The first integral in the right side of Eq. (14-80) is by definition the initial voltage $v(t_c)$. The second integral can be easily integrated to obtain a linear function of t because $i(\tau)$ is constant for any horizontal segment. We have shown, therefore, that the voltage waveform corresponding to any horizontal segment is always a linear "ramp" and can be expressed as

$$v(t) = v(t_c) - \frac{i}{C}(t - t_c) \qquad (14\text{-}81)$$

where i is the current corresponding to the horizontal segment. From Eq. (14-81) it is easy to compute the time it takes the linear

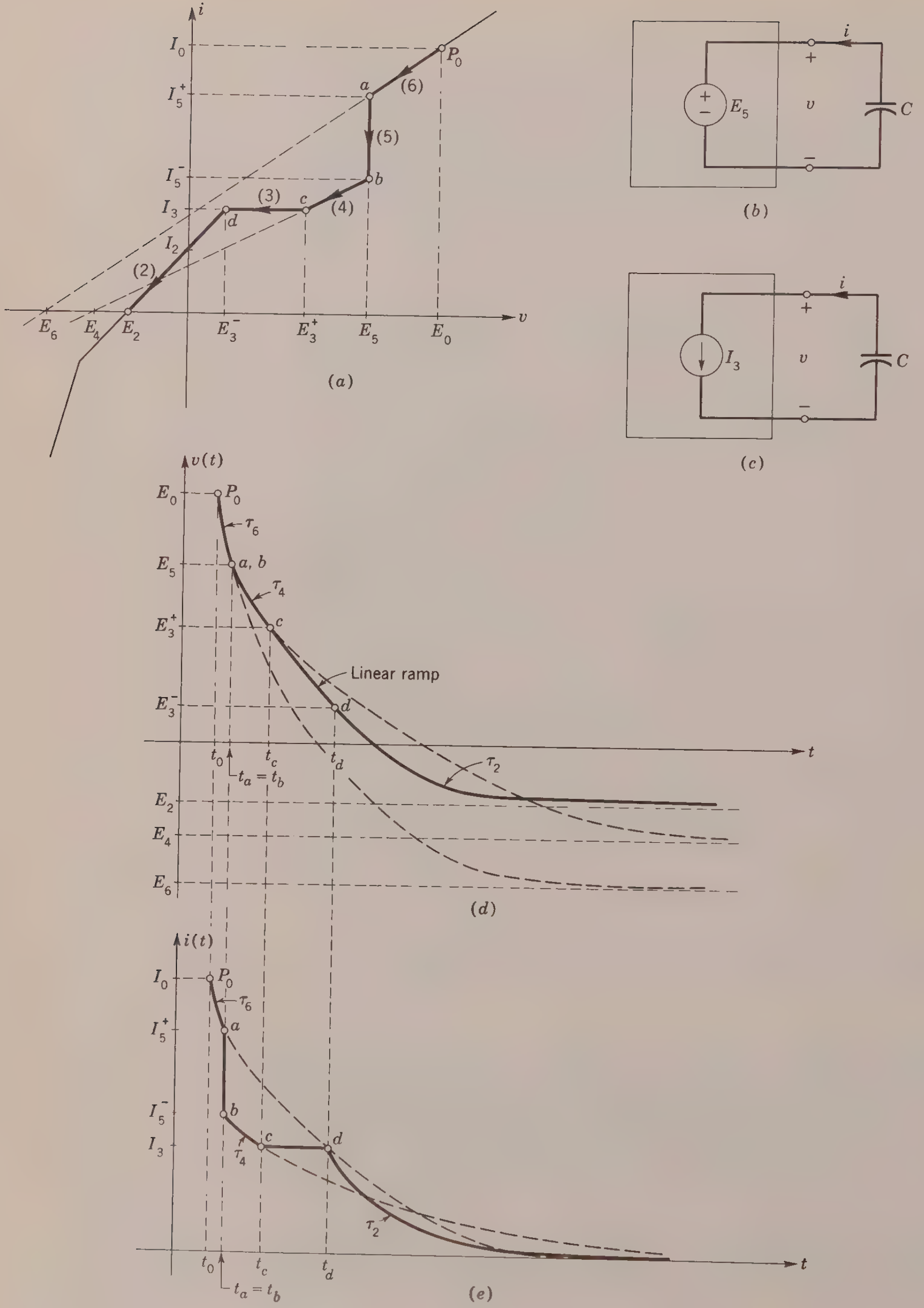


Fig. 14-29. A first-order network with a DP plot containing both horizontal and vertical segments and its associated solution waveforms.

ramp $v(t)$ to move from $v(t_0)$ to any point $v(t)$; namely,

$$(t - t_c) = \frac{-C[v(t) - v(t_c)]}{i} \tag{14-82}$$

If we let $\Delta T \equiv (t - t_c)$ and $\Delta v \equiv [v(t) - v(t_c)]$, Eq. (14-82) becomes

$$\Delta T = -C\frac{\Delta v}{i} \tag{14-83}$$

This equation could have been derived more easily upon replacing the differentials in the expression $i = -C\, dv/dt$ by incremental quantities, a legitimate operation since $v(t)$ is linear.

Applying Eqs. (14-81) and (14-82) to segment 3 of the present DP plot, we obtain

$$v(t) = E_3^+ - \frac{I_3}{C}(t - t_c) \tag{14-84}$$

and

$$t_d = t_c - C\frac{(E_3^- - E_3^+)}{I_3} \tag{14-85}$$

where t_c and t_d denote the time when the dynamic route arrives at points c and d, respectively.

The complete waveform $v(t)$ is, therefore, as shown in Fig. 14-29*d*. The corresponding current $i(t)$ can be obtained either from $i = -C\, dv/dt$ or from the dynamic route as before. The result is shown in Fig. 14-29*e*.

By a dual procedure it is just as easy to determine the waveforms $v(t)$ and $i(t)$ when the energy-storage element is a linear inductor. In this case, the current $i(t)$ will become a linear ramp along any *vertical* segment and will jump instantaneously on a horizontal segment. The equation for $i(t)$ corresponding to a vertical segment starting at $i(t_0)$ is simply the dual of Eq. (14-81); namely,

$$i(t) = i(t_0) - \frac{v}{L}(t - t_0) \tag{14-86}$$

Similarly, the time interval from the initial state to any point on the vertical segment is given by

$$t - t_0 = -\frac{L[i(t) - i(t_0)]}{v} \tag{14-87}$$

or

$$\Delta T = -L\frac{\Delta i}{v} \tag{14-88}$$

where

$\Delta T \equiv (t - t_0)$ and $\Delta i \equiv i(t) - i(t_0)$.

14-8-3 NETWORKS CONTANING A PIECEWISE-LINEAR ENERGY-STORAGE ELEMENT ACROSS A NONLINEAR RESISTIVE NETWORK CHARACTERIZED BY PIECEWISE-LINEAR SEGMENTS WITH ARBITRARY SLOPES

Let us now consider the more general case where the energy-storage element is also nonlinear. To be specific, we shall assume that the capacitor is characterized by a piecewise-linear v-q curve and that the inductor is characterized by a piecewise-linear i-φ curve. Therefore each segment of the v-q curve can be described by

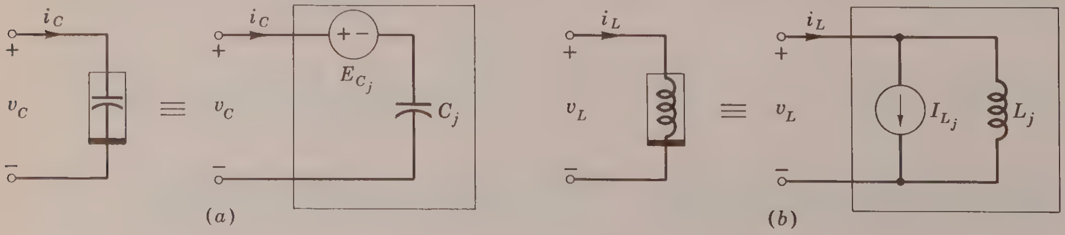
$$v_C = \frac{1}{C_j}q + E_{C_j} \qquad j = 1, \ldots, n \tag{14-89}$$

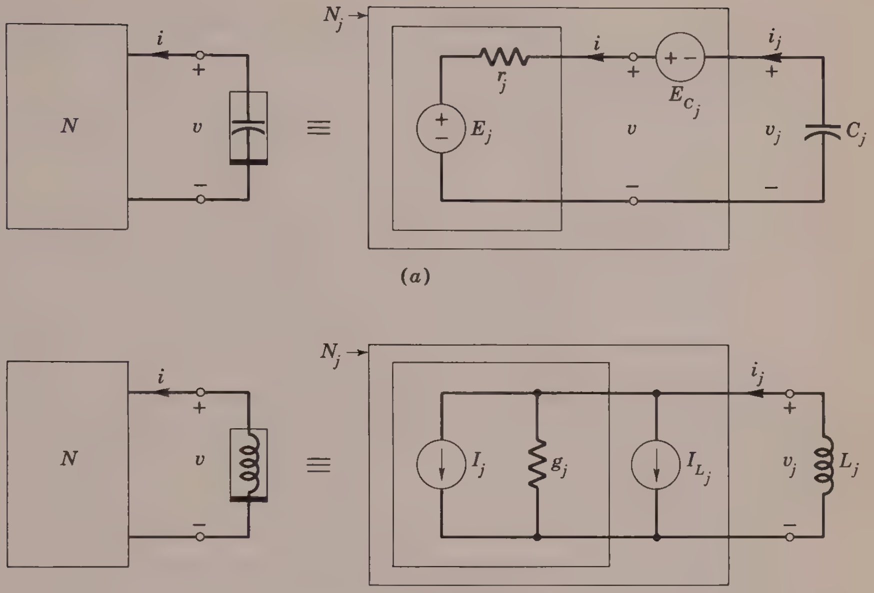
Similarly, each segment of the i-φ curve can be described by

$$i_L = \frac{1}{L_j}\varphi + I_{L_j} \qquad j = 1, \ldots, n \tag{14-90}$$

Equations (14-89) and (14-90) can be represented by the iterative equivalent linear circuits shown in Fig. 14-30*a* and *b*, respectively. For the nonlinear capacitor we can replace the network by the iterative linear circuit shown in Fig. 14-31*a*. Observe that we can combine the voltage source E_{C_j} with the original network N to obtain an equivalent resistive subnetwork N_j connected with a *linear*

Fig. 14-30. A piecewise-linear capacitor can be replaced by a voltage source and a linear capacitor in series. A piecewise-linear inductor can be replaced by a current source and a linear inductor in parallel.





(b)

Fig. 14-31. An arbitrary first-order nonlinear autonomous network containing a piecewise-linear capacitor or inductor can always be converted into the equivalent forms shown in (*a*) or (*b*).

capacitor C_j. Similarly, for the nonlinear inductor we can replace the network by the iterative linear circuit shown in Fig. 14-31*b*. Observe that we can combine the current source I_{L_j} with the original network N to obtain an equivalent resistive subnetwork N_j connected with a *linear* inductor L_j. In either case, the resulting networks shown in Fig. 14-31 reduce to the same forms as those considered previously in Secs. 14-8-1 and 14-8-2.

From the above observation it is clear that the same methods discussed previously are applicable here. But this time we have

$$i_j = -C_j \frac{dv_j}{dt} \tag{14-91}$$

for the capacitor and

$$v_j = -L_j \frac{di_j}{dt}$$

for the inductor. The time constants will now depend on both the resistance of each segment of the DP plot and the capacitance or inductance of the corresponding segment of the v-q or i-φ curve. In order to keep track of which segment combination to use at any

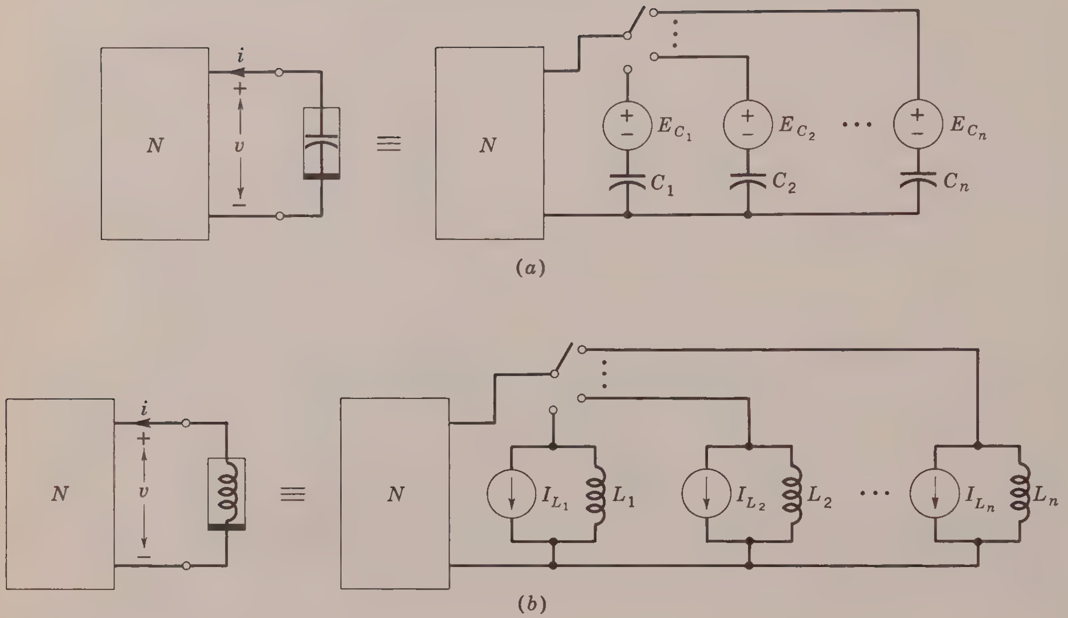


Fig. 14-32. The piecewise-linear networks in Fig. 14-31 can be transformed into an equivalent network containing switches.

given time, it is sometimes convenient to redraw the networks in Fig. 14-31 into the more explicit form containing switches, as shown in Fig. 14-32. Observe that, in this case, the DP plot seen by the linear capacitor C_j or inductor L_j changes whenever the terminal voltage v or current i of the original resistive network N reaches a breakpoint of the v-q curve or the i-φ curve, respectively. A network with this property is an example of a first-order switching network to be discussed in the next chapter. Therefore, we should be able to handle this more general class of nonlinear networks after studying Chap. 15.

Exercise 1: Describe the step-by-step procedure for obtaining the capacitor current waveform or the inductor voltage waveform of any first-order autonomous network by *inspection.*

Exercise 2: Replace the capacitor in Fig. 14-16*a* by an inductor and sketch the inductor voltage and current waveforms corresponding to an initial current $i(t_0) = I_0$.

Exercise 3: Replace the inductor in Fig. 14-22*a* by a capacitor and sketch the capacitor current and voltage waveforms corresponding to an initial voltage $v(t_0) = E_0$ (assume N is operating initially on segment 4).

Exercise 4: Specify the DP plot of the network N in Fig. 14-28*a* in order to generate the dual waveforms of Fig. 14-28*c* and *d*, if the inductor is replaced by a capacitor.

Exercise 5: Specify the DP plot of a network so that when connected across a 1-μF capacitor, a 1-kHz, 10-volt symmetrical triangular voltage waveform is generated.

14-9 SUMMARY

Two basic approaches: qualitative and quantitative The first approach finds the equilibrium states and their associated trajectories. It gives useful qualitative information. This approach is tedious if the solution waveforms are to be found because of the necessity for time scaling. The second approach is the piecewise-linear method and can be used to find the solution waveforms by inspection. These two approaches supplement each other because the former is qualitative and the latter is quantitative in nature.

General stability criteria An equilibrium state P_j of the normal-form equation $dx/dt = f(x)$ is stable if there exists an interval around P_j such that $f(x)$ is a monotonically decreasing function of x. Otherwise, it is unstable.

Corollary to the stability criteria If the slope of $f(x)$ is nonzero at *each* equilibrium state P_j, then P_j is stable if the slope of $f(x)$ at P_j is negative. Otherwise, P_j is unstable. Moreover, if $f(x)$ is a single-valued function, then the singular points must alternate between stability and instability.

Stability criteria for two common circuit configurations For a capacitor circuit in configuration 1, an equilibrium state P_j is stable if, and only if, $1/R > -g'(v)$, where $g'(v)$ is the incremental conductance evaluated at the equilibrium state P_j. For an inductor circuit in configuration 2, an equilibrium state P_j is stable if, and only if, $R > -r'(i)$, where $r'(i)$ is the incremental resistance evaluated at the equilibrium state P_j.

Basic principle for handling multivalued functions When $f(x)$ of the normal-form equation is not a single-valued function, two techniques are available for obtaining a unique solution:

1. Make use of the inertia postulate and the principles of conservation of charge and flux linkages. The solution waveforms in this case will contain jump discontinuities.

2. Introduce additional parasitic elements at strategic locations so that the normal-form equations can be obtained. The existence of the normal-form equations precludes the possibility of any jump phenomenon.

Piecewise-linear method The first step is to draw the dynamic route corresponding to a given initial state. Each segment of the dynamic route gives rise to an exponential waveform whose exact form can be found by determining, by inspection, three pieces of information, namely, the initial state, the equilibrium state, and the time constant associated with the segment. The pertinent equations are given in Table 14-1. When a segment is either horizontal or vertical, the corresponding solution waveform is linear and can also be determined by inspection.

PROBLEMS

14-1 Let $x = x_e$ be an equilibrium state of the system $dx/dt = f(x)$.

(a) Prove that the equilibrium state x_e is stable if, and only if, $(x - x_e)f(x) < 0$ for all values of x in the vicinity of x_e; namely, $(x_e - \epsilon) < x < (x_e + \epsilon)$, where ϵ is any positive number which does not cause the initial state to move beyond any adjacent equilibrium state.

(b) Apply the above criterion to test the stability of each of the equilibrium states of the equation $dx/dt = \cosh x - 10$.

14-2 Consider the two first-order circuits shown in Fig. P14-2*a* and *b*. The thermistor is assumed to be operating at $T = 50°\text{F}$.

(a) Find the operating points of these networks when the capacitor is open-circuited and the inductor is short-circuited.

(b) For the tunnel-diode circuit shown in Fig. P14-2*a*, find the operating-point voltage that will be measured in steady state corresponding to each of the initial states $v_C(0) = 0.25$, 0.1, and 0.5 volt.

(c) For the thermistor circuit shown in Fig. P14-2*b*, find the operating-point current that will be measured in steady state corresponding to each of the initial states $i_L(0) = 5$, 25, 50 ma.

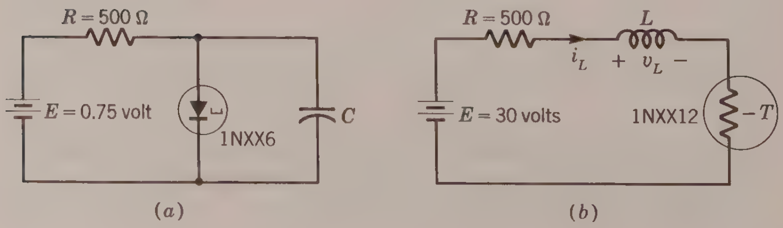


Fig. P14-2.

(d) In theory, is it possible to operate at the middle equilibrium state? If so, specify the initial state for achieving this goal. Is this operating point observable in practice? Explain why.

(e) Specify the range of permissible initial states for which the circuit will be operating at a given equilibrium state.

14-3 Consider the network shown in Fig. P14-3a. The DP plot of the resistive subnetwork N is shown in Fig. P14-3b, and the v-q curve of the nonlinear capacitor is shown in Fig. P14-3c.

(a) Write the equations of motion in terms of the state variable q. Plot the associated function $f(q)$.

(b) Find all equilibrium states and determine their stability.

(c) Assuming an initial charge $q(0) = -1$ coul, find $q(t)$ analytically.

(d) Repeat part (c) by the qualitative method and sketch the trajectory accurately.

(e) Determine the time scale analytically. Repeat by applying the trapezoidal rule.

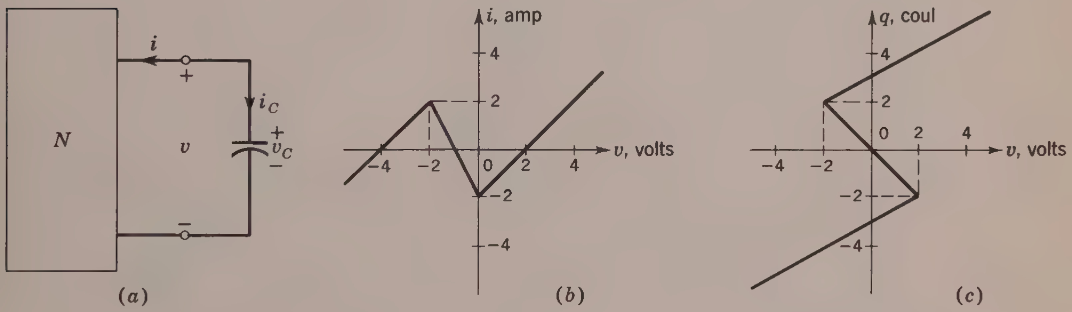
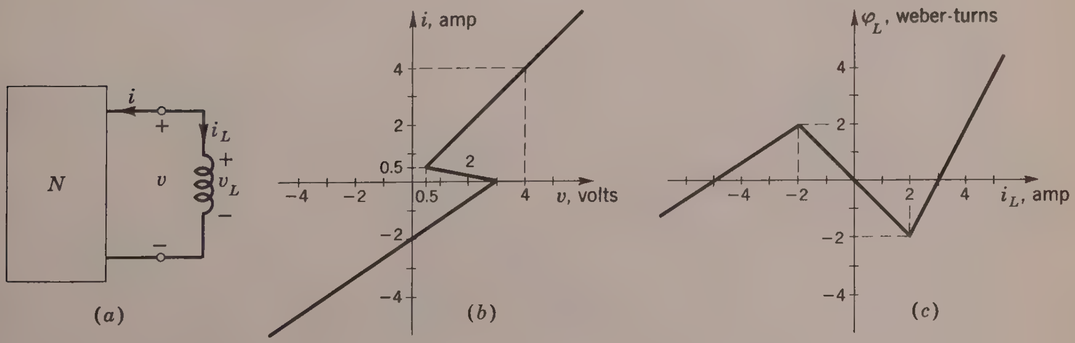


Fig. P14-3.

14-4 Repeat Prob. 14-3 for the network shown in Fig. P14-4. Assume that the initial inductor current $i_L(0) = 1.8$ amp and choose the inductor flux linkage φ_L to be the state variable.

Fig. P14-4.



14-5 Consider the first-order network shown in Fig. P14-5, where the DP plot of the black box N is given by $i_1 = v_1^3$ and the v-q curve of the capacitor is given by $q = v^3$.

(a) If the initial charge $q(0) = 1$ coul, find $q(t)$ analytically.

(b) Repeat (*a*) by the qualitative method.

(c) Find $v(t)$ and $i(t)$ corresponding to the same initial state $q(0) = 1$ coul.

(d) Determine $v(t)$ directly from $q(t)$ and the v-q curve.

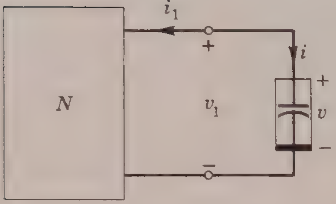


Fig. P14-5.

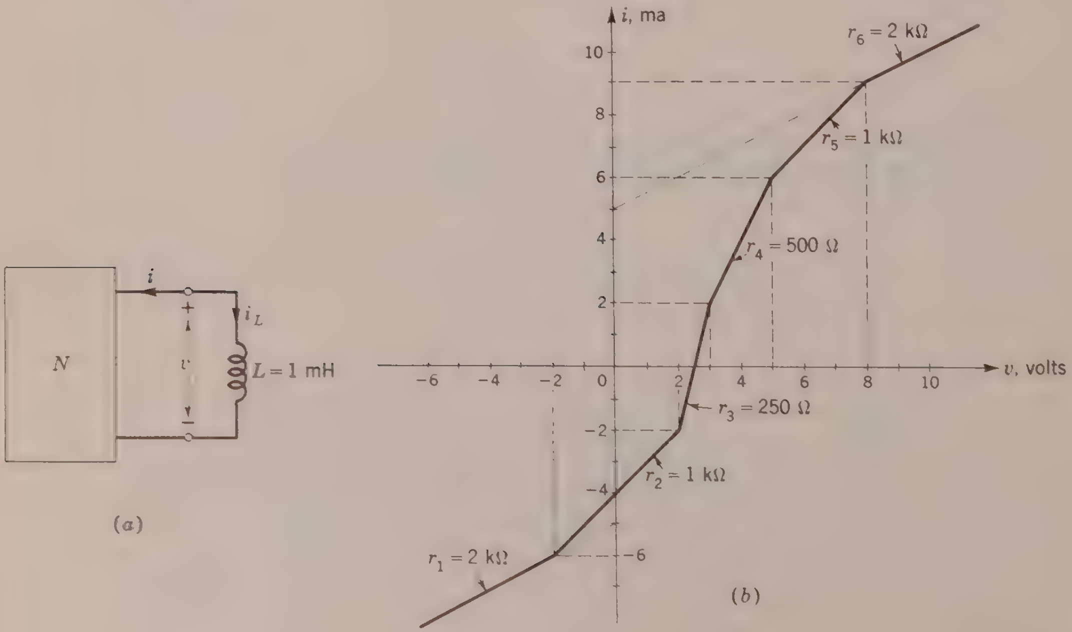
14-6 Consider the network shown in Fig. P14-5 but with the DP plot given by $i_1 = (v_1 + \frac{1}{16}v_1^3)10^{-6}$ amp.

(a) Assuming an initial voltage $v_C(0) = 5$ volts and a linear capacitor $C = 1\ \mu\text{F}$, find $v(t)$ analytically.

(b) Repeat (*a*) by applying the qualitative method and the trapezoidal rule.

14-7 Consider the first-order network shown in Fig. P14-7*a*, where the DP plot of N is given in Fig. P14-7*b*. Find the inductor current $i_L(t)$ and inductor voltage $v_L(t)$ for $t \geq 0$ corresponding to the

Fig. P14-7.



initial states $i_L(0) = 7$ ma and -10 ma. Sketch the time functions accurately and label all pertinent quantities.

14-8 Consider the network shown in Fig. P14-8*a* where the DP plot of N is given by Fig. P14-8*b*. If the initial charge in the capacitor is $q(0) = -12 \times 10^{-6}$ coul, find $i_C(t)$ and $v_C(t)$ for $t \geq 0$. Sketch the time functions accurately and label all pertinent quantities.

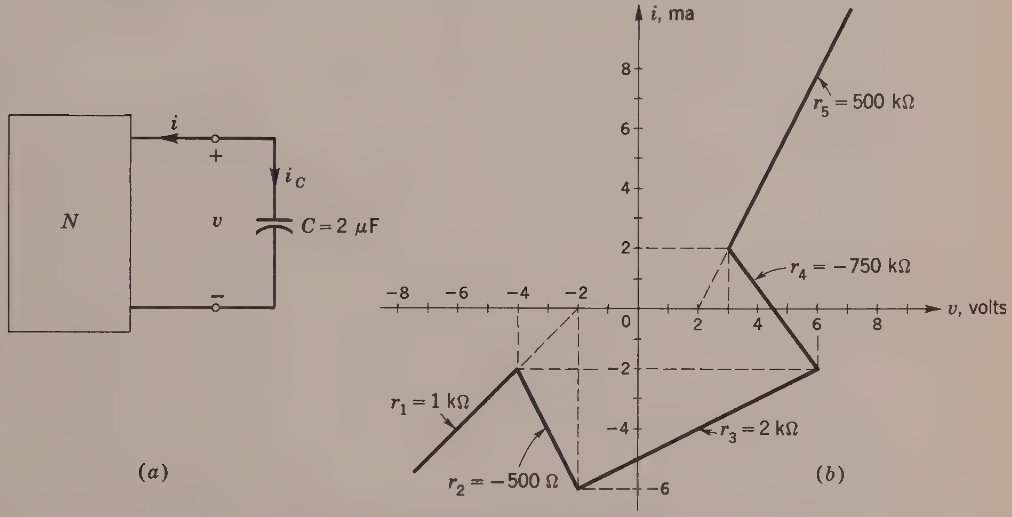


Fig. P14-8.

14-9 Consider the network shown in Fig. P14-9*a* where the DP plot of N is given by Fig. P14-9*b*. Find $i_L(t)$ and $v_L(t)$ for $t \geq 2$ sec corresponding to the initial current $i_L(2) = 5$ ma. Sketch the time functions accurately and label all pertinent quantities. Assume that any jump in the dynamic route is toward the nearest positive-slope segment of the DP plot.

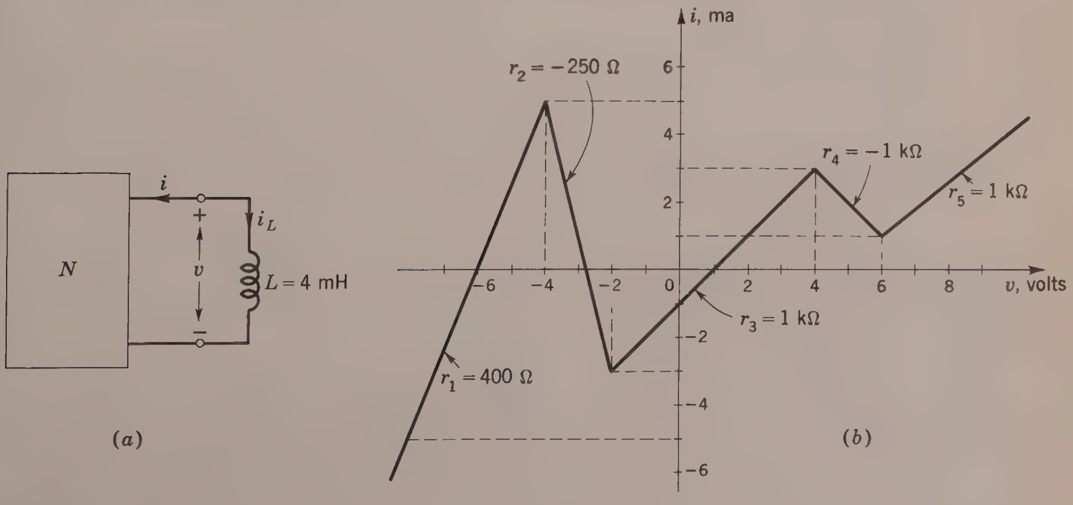


Fig. P14-9.

14-10 Assuming an initial charge $q(0) = 10^{-6}$ coul, find the output voltage $v_o(t)$ of the network shown in Fig. P14-10. Sketch the waveforms accurately and label all pertinent quantities.

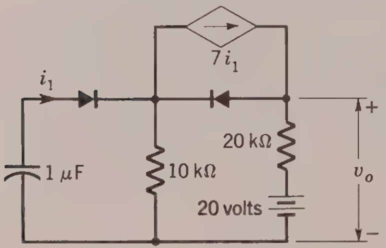


Fig. P14-10.

14-11 Consider the two zener-diode circuits shown in Fig. P14-11*a* and *b*. Let the zener diodes be represented by the usual piecewise-linear *v-i* curve with a breakdown voltage E_z.

(a) Find the values of E_z and R in the circuit of Fig. P14-11*a* in order to generate a 20-volt peak-to-peak 10-kHz square wave $v_o(t)$.

(b) Assuming that $E_z = 5$ volts and $i_L(0) = 0.5$ ma in the circuit shown in Fig. P14-11*b*, find $v_L(t)$, $i_L(t)$, and their periods of oscillation.

(c) Repeat (*b*) with $R_1 = 0$.

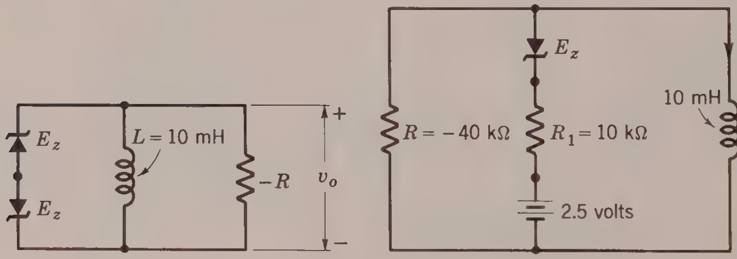


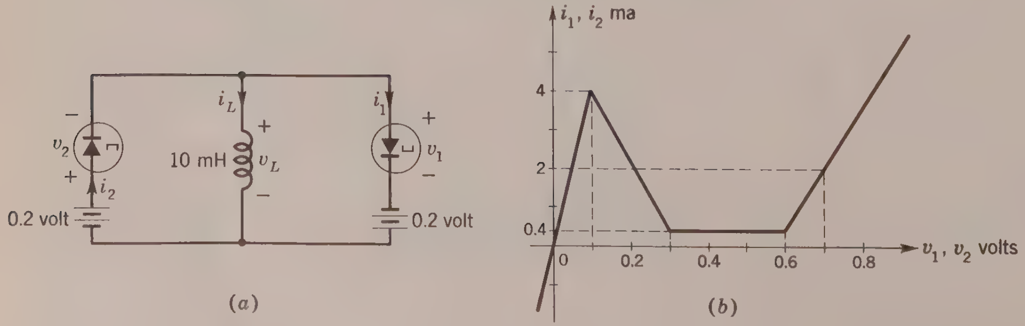
Fig. P14-11. (*a*) (*b*)

14-12 Consider the tunnel-diode circuit shown in Fig. P14-12*a* and the tunnel-diode *v-i* curve shown in Fig. P14-12*b*.

(a) If the initial inductor current is $i_L(0) = 5$ ma, find $i_L(t)$ and $v_L(t)$ for $t \geq 0$. Sketch all waveforms accurately.

(b) Find the frequency of oscillation.

Fig. P14-12.

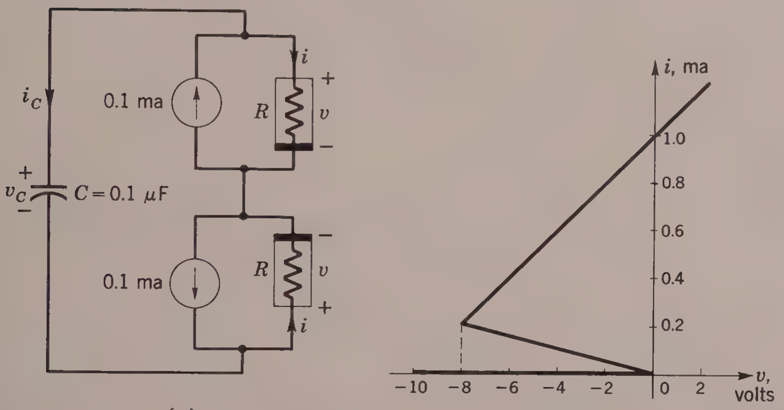


(*a*) (*b*)

14-13 Consider the networks shown in Fig. P14-13*a* where the *v*-*i* curve of the nonlinear resistor *R* is shown in Fig. P14-13*b*.

(a) If the voltage across the capacitor at $t = 10$ msec is 10 volts, find $i_C(t)$ and $v_C(t)$ for $t \geq 10$ msec. Sketch all waveforms accurately.

(b) Find the value of C so that the frequency of steady-state oscillation is equal to 1 kHz.



(*a*) (*b*)

Fig. P14-13.

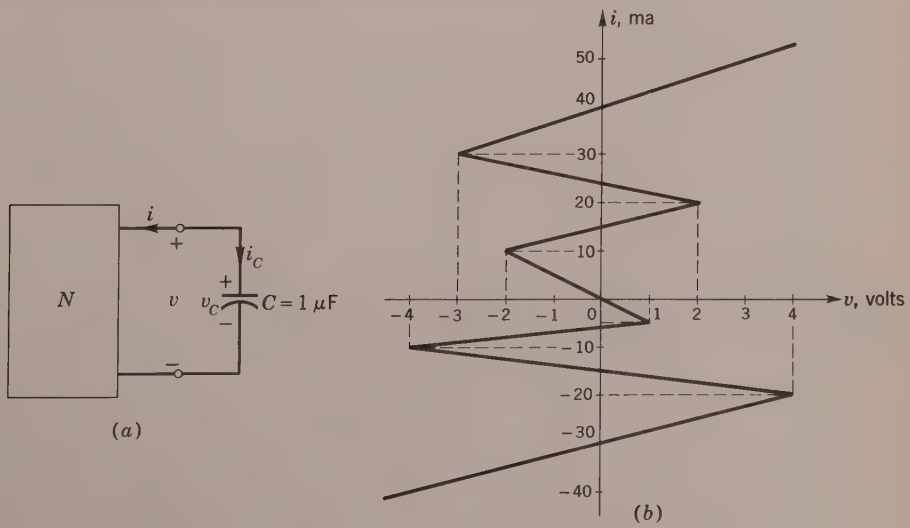
14-14 Consider the network shown in Fig. P14-14*a*, where the DP plot of the black box *N* is shown in Fig. P14-14*b*.

(a) Assuming an initial voltage $v_C(0) = -5$ volts, find the current $i_C(t)$ and the voltage $v_C(t)$ for $t \geq 0$. Sketch all waveforms accurately.

(b) After the trajectory has reached a steady state, $i_C(t)$ and $v_C(t)$ become periodic. Find the frequency of oscillation.

NOTE: If there is more than one possible segment for the dynamic route to jump, choose the nearest segment with a positive slope.

Fig. P14-14.



(*a*) (*b*)

14-15 Consider the network shown in Fig. P14-15*a*, where the DP plot of N is given in Fig. P14-15*b*.

(a) If the initial voltage $v_C(0) = -10$ volts, find $i_C(t)$ and $v_C(t)$ for $t \geq 0$. Sketch the time functions accurately and label all pertinent quantities.

(b) Find the frequency of oscillation in the steady state.

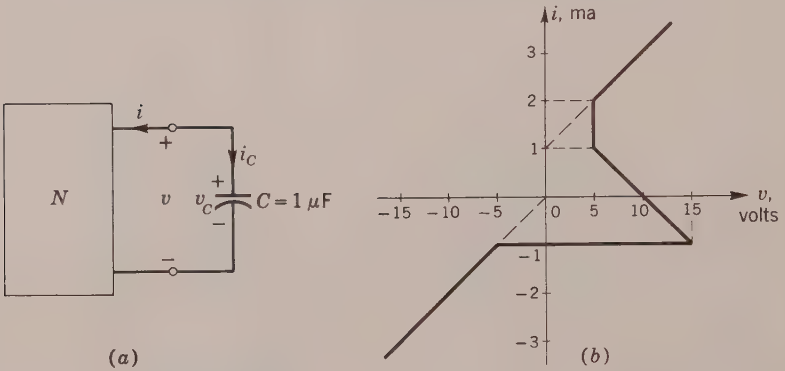


Fig. P14-15. (*a*) (*b*)

14-16 Consider the network shown in Fig. P14-16*a*, where the DP plot of N is given in Fig. P14-16*b*. If the initial current $i_L(0) = 3.0$ ma, find $i_L(t)$ and $v_L(t)$ for $t \geq 0$. Sketch all waveforms accurately.

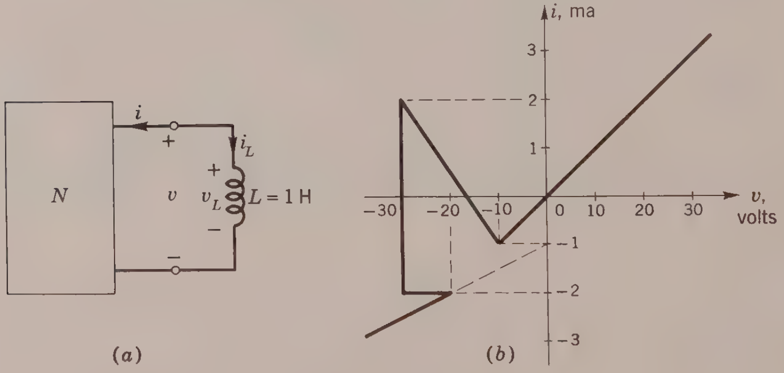


Fig. P14-16. (*a*) (*b*)

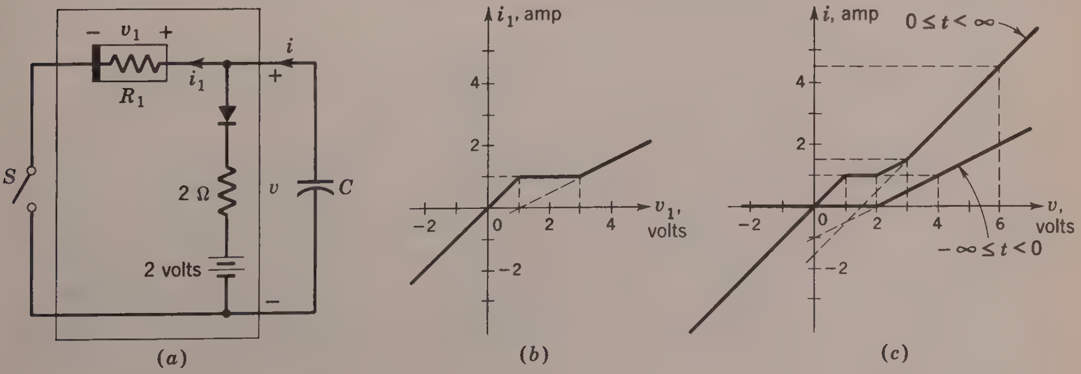
15 ANALYSIS OF FIRST-ORDER NONLINEAR SWITCHING NETWORKS

15-1 WHAT IS A FIRST-ORDER SWITCHING NETWORK?

The principles presented in the preceding chapter are generally not applicable to nonautonomous networks; the reason is that, unlike the autonomous case, the DP plot associated with a nonautonomous network does not remain fixed but changes with time. However, there is an important subclass of nonautonomous networks which behaves like autonomous networks during some specified intervals of time. This class of networks can be distinguished by the presence of one or more switches (mechanical or electronic) or stepwise switching signals, or both, and are henceforth called *switching networks.*

An example of a switching network containing a switch is shown in Fig. 15-1*a*. If the v_1-i_1 curve of R_1 is as shown in Fig. 15-1*b*, and if the switch S is off for $-\infty \leq t < 0$ and on for all times $t \geq 0$, then the DP plots of the network during the corresponding time intervals are shown in Fig. 15-1*c*. Another example

Fig. 15-1. An example of a first-order switching network whose DP plot depends on whether the switch is on or off.



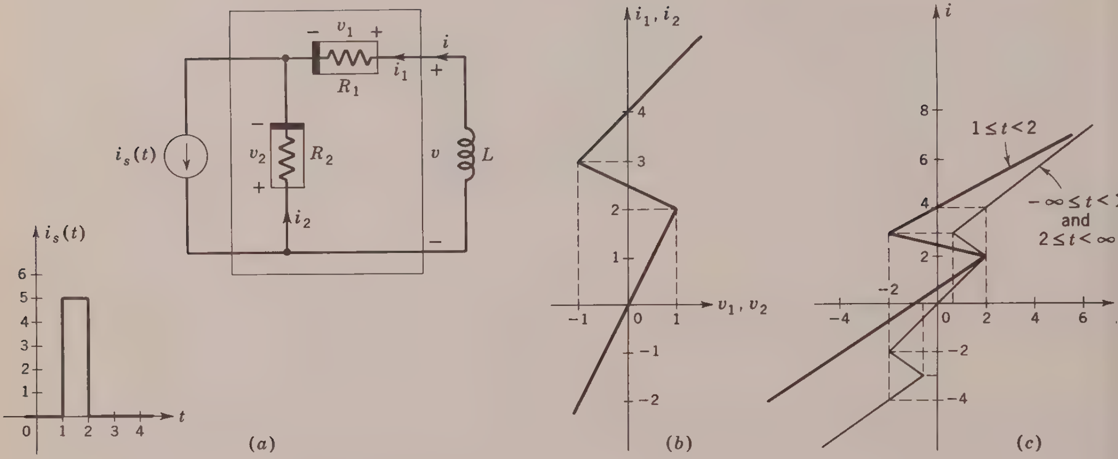


Fig. 15-2. An example of a first-order switching network whose DP plot changes according to a stepwise switching signal.

of a switching network whose DP plot changes according to a stepwise switching signal is shown in Fig. 15-2*a*. The two nonlinear resistors are characterized by the same v_j-i_j curve shown in Fig. 15-2*b*. The DP plot during the time intervals $-\infty \le t < 1$ and $2 \le t < \infty$ is easily found to be given by the lower v-i curve shown in Fig. 15-2*c*. The corresponding DP plot during $1 \le t < 2$ is given in the same figure by the upper v-i curve.

More complicated switching networks can be found where both switches and stepwise switching signals are present. In any case, however, it is possible to partition the time axis into a finite number of intervals such that during each interval, the energy-storage element "sees" a fixed DP plot. We shall now extend the techniques in the preceding chapter to analyze this class of nonautonomous networks.

15-2 ANALYSIS OF FIRST-ORDER LINEAR SWITCHING NETWORKS

A switching network containing only linear resistors and one linear inductor or one linear capacitor is called a *first-order linear switching network*. This class of networks is important because: (1) there are many practical switching networks which are of this type, and (2) any first-order *nonlinear* switching network can be analyzed as a number of linear switching networks with appropriate initial and transition states. Since basically no new concept is involved, we shall consider only examples to illustrate the details.

EXAMPLES

1. Consider the linear switching network shown in Fig. 15-3*a*. The switch S has been "open" for a "long time." By this we mean a long time compared with the largest time constant of the network so that for all practical purposes, the network can be assumed to be operating in its equilibrium state. If the switch is closed at $t = 1$ sec and then "reopened" at $t = 2$ sec, find the output voltage $v_o(t)$ for all times $t \geq 0$.

   Since the network is linear, we can replace the resistive subnetwork across the capacitor by its Thévenin equivalent circuit. This is shown in Fig. 15-3*b* for the case when S is open and in Fig. 15-3*c* when S is closed. The corresponding DP plots are shown by the upper and the lower straight lines in Fig. 15-3*d*.

   Since the network is assumed to have reached its equilibrium state, it is operating at point a of the upper DP plot during the time interval $0 \leq t < 1$. At $t = 1$ sec, the DP plot switches to the lower line. Since the capacitor voltage cannot change instantaneously, the transition state at $t = 1^+$ must be at point $b(v = 6)$ in the lower half-plane. Accordingly, the dynamic route will move up along the lower DP plot until it reaches some point c corresponding to $t = 2$ sec when the DP plot switches back to the upper line. The transition state at $t = 2^+$ sec must be at point d (directly above c) in the upper half-plane, and hence the dynamic route must move down along the upper DP plot until it arrives at its equilibrium state again. The complete dynamic route is shown by the solid lines in Fig. 15-3*d*.

   Applying the principles in the preceding chapter, we can immediately draw the solution waveform $v_C(t)$ as shown in Fig. 15-3*e* for all times $t \geq 0$. The corresponding equations are given by

$$v_C(t) = \begin{cases} 10 - 4e^{-[(t-1)/0.9]} & 1 \leq t < 2 \\ 6 + 2.68e^{-[(t-2)/1.0]} & 2 \leq t < 8 \end{cases} \tag{15-1}$$

   The voltage $v_o(t)$ can now be obtained by inspection of Fig. 15-3*a*; namely,

$$v_o(t) = \begin{cases} R_3 i_C = R_3 C \dfrac{dv_C}{dt} & \\ 1.78e^{-[(t-1)/0.9]} & 1 \leq t < 2 \\ -1.07e^{-[(t-2)/1.0]} & 2 \leq t < \infty \end{cases} \tag{15-2}$$

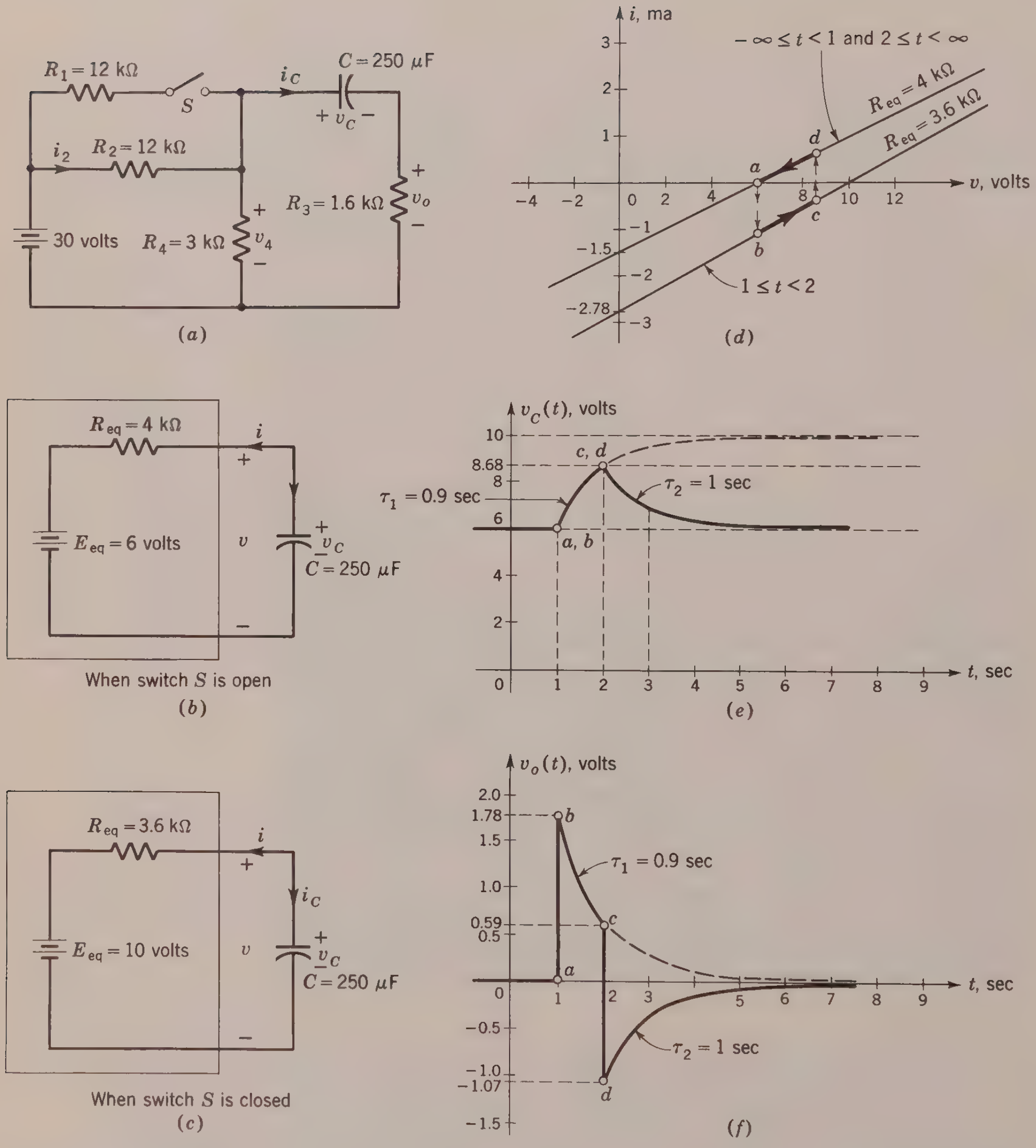


Fig. 15-3. An example illustrating the steps involved in the solution of a typical linear switching network.

The resulting waveform for $v_o(t)$ is sketched in Fig. 15-3*f* and is seen to consist of exponential segments with time constants identical with $v_C(t)$ during the corresponding intervals of time. Indeed, by applying the voltage-source substitution theorem, we can replace the capacitor by a voltage source $v_C(t)$ having the waveform shown in Fig. 15-3*e*. The resulting voltages and currents in all resistors must, therefore, consist also of exponential segments with identical time constants for $v_C(t)$ during corresponding intervals of time. This follows because the waveform solutions of any resistive network are always identical with the waveform of the excitation function. This observation is important enough to be summarized as a theorem.

EXPONENTIAL WAVEFORM SOLUTION THEOREM

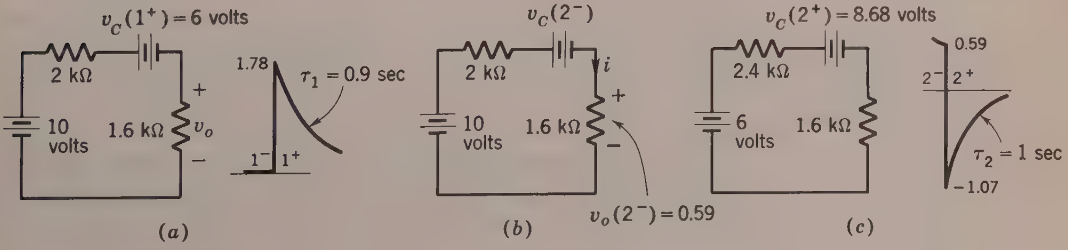
The solution waveform of the voltage across any terminal pair or the current in any branch of a first-order linear switching network always consists of a sequence of exponential waveforms $k_i e^{-[(t-t_i)/\tau_i]}$ during each time interval (t_i, t_{i+1}). The same time constant τ_i applies to any voltage and current solution during each time interval.

Applying this theorem, we can determine the voltage or current waveform of any resistor of a linear switching network by a more direct method. All we need to find are the three pertinent pieces of information describing an exponential. For example, consider the network shown in Fig. 15-3*a* again. Since the capacitor current is zero when the network is in equilibrium, we have $v_o(t) = 0$ for $t < 1$; hence $v_o(1^-) = 0$. To determine $v_o(1^+)$, we observe that at $t = 1^+$ sec, the capacitor voltage can be replaced by a battery $v_C = 6$ volts as shown in Fig. 15-4*a*. (The 2-kΩ resistor and 10-volt battery represent the Thévenin equivalent of the resistive network to the left of the capacitor with the switch closed.) The voltage $v_o(t)$ at $t = 1^+$ is, therefore, equal to

$$v_o(1^+) = 1.78 \text{ volts} \tag{15-3}$$

The time constant is identical with that of $v_C(t)$ during $1 \leq t < 2$ and is given by $\tau_1 = 0.9$ sec. The equilibrium state is given by

Fig. 15-4. The three instantaneous networks needed to determine the transition states.



(*a*) (*b*) (*c*)

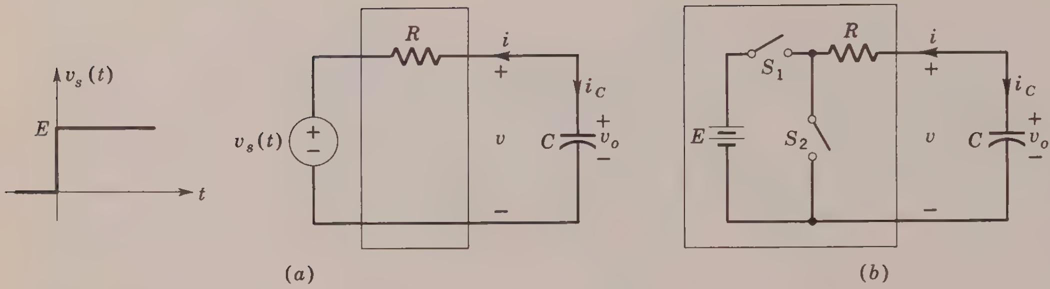


Fig. 15-5. A simple linear switching network containing a stepwise switching signal and its equivalent network.

$v_o(t_e) = 0$. From this information we obtain the same exponential waveform during $1 \leq t < 2$ as shown in Fig. 15-3*f*.

At $t = 2$ sec, the switch S is reopened, and hence it is necessary to determine the three pieces of information $v_o(2^+)$, τ_2, and $v_o(t_e)$. Clearly, $\tau_2 = 1$ and $v_o(t_e) = 0$. It remains to find $v_o(2^+)$. Since the voltage across a resistor may jump instantaneously, $v_o(2^+)$ need not equal $v_o(2^-)$. However, we do know that $v_C(2^+) = v_C(2^-)$, and from this we can find $v_o(2^+)$. To show how this is obtained, let us determine $v_C(2^-)$ first. The equivalent circuit that applies at $t = 2^-$ is shown in Fig. 15-4*b*. Since $v_o(2^-) = 0.59$, $i_C(2^-) = 0.59/1.6k = 0.37$ ma, and

$$v_C(2^-) = 10 - (2)(0.37) - 0.59 = 8.68 \tag{15-4}$$

Therefore, $v_C(2^+) = 8.68$, and the equivalent circuit that applies at $t = 2^+$ is shown in Fig. 15-4*c*. The capacitor is again replaced by a battery, and the resistive network across the capacitor is replaced by the Thévenin equivalent with the switch opened. We can now compute for $v_o(2^+)$ as follows:

$$v_o(2^+) = \frac{1.6}{4.0}(-2.68) = -1.07 \tag{15-5}$$

The final solution for $t \geq 2$ can now be obtained as shown in Fig. 15-3*f*.

Observe that the above alternate procedure is a direct one because we have solved for $v_o(t)$ without having to solve for $v_C(t)$ first. For this example the amount of work involved in the two methods is about the same. In other networks one method might be simpler than the other. However, in all cases the first method is conceptually simpler and offers less chance for error.

2. Consider the simple R-C switching network shown in Fig. 15-5*a*. The stepwise switching signal $v_s(t)$ is equal to zero for $t < 0$ and is equal to E for $t \geq 0$. In spite of its simplicity, this network is very important because it represents a large class of

practical networks. For example, the resistive subnetwork might represent the equivalent circuit of a square-wave voltage generator, and the capacitance C might represent the "parasitic" capacitance between the output terminals. It is of utmost importance to know the effect of this parasitic element on the output waveform.

For purposes of analysis a switching signal can always be represented by an appropriate combination of batteries and switches. For example, the switching signal of Fig. 15-5*a* can be replaced by the equivalent switching circuit shown in Fig. 15-5*b*. In this case, S_1 is open while S_2 is closed for $t < 0$. This will simulate the condition $v_s(t) = 0$ for $t < 0$. Similarly, S_1 is closed while S_2 is open for $t \geq 0$ to simulate the condition $v_s(t) = E$.

To analyze this network, observe first that since S_2 is closed for $t < 0$, $v_o(0^-) = 0$. The three pertinent pieces of information for drawing the exponential are easily obtained by inspection; namely, $v_o(0^+) = v_o(0^-) = 0$, $\tau = RC$, and $v_o(t_e) = E$. The corresponding solution is

$$v_o(t) = E - Ee^{-(t/RC)} \qquad t \geq 0 \tag{15-6}$$

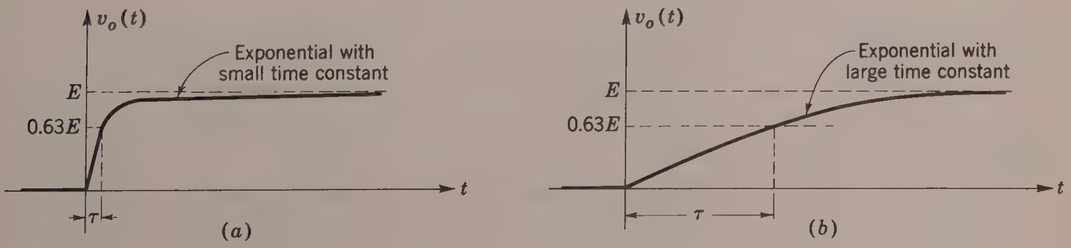
This waveform is shown in Fig. 15-6*a* and *b* illustrating the relative difference between a large and a small value of τ.

Observe that in the limit where $\tau = 0$, Eq. 15-6 reduces to

$$v_o(t) = E \qquad t \geq 0 \tag{15-7}$$

as expected, since this corresponds to the case where no parasitic element is present. The effect of the parasitic capacitance is, therefore, to distort the actual output waveform from the ideal waveform given by Eq. (15-7). Clearly, the larger τ is, the greater the distortion. It is highly desirable, therefore, to reduce τ as much as possible in order to obtain a faithful reproduction of the input signal. In practice, this is accomplished by careful wiring and by using as short leads as possible so as to minimize the parasitic capacitance.

Fig. 15-6. Two typical exponential waveform solutions, (*a*) with a small time constant and (*b*) with a large time constant.



From this example it is seen that the time constant τ is a useful parameter not only because it is needed to draw the exponential waveform accurately, but also because it gives a measure of how "good" a system is in reproducing an input signal. Unfortunately, the concept of a time constant is meaningful only for first-order networks. Since it is desirable to have a measure of distortion for higher-order networks as well, it is standard practice to define another parameter, t_r, known as the *rise time* of the network. The rise time is the time it takes the output waveform to rise from 10 percent of its final value at equilibrium to 90 percent of its final value when the input is a stepwise signal. For first-order networks a simple relationship between the rise time t_r and the time constant τ of a network can be easily derived, namely,

$$t_r = \tau \ln \frac{E - 0.1E}{E - 0.9E} = \tau \ln 9 = 2.2\,\tau \tag{15-8}$$

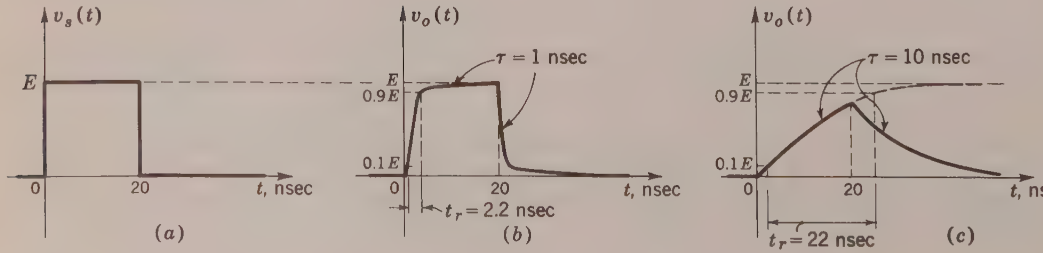
In order to appreciate the significance of the rise time t_r, let the network shown in Fig. 15-5*a* be excited by the rectangular pulse shown in Fig. 15-7*a*. Assuming that $R = 1\ \text{k}\Omega$ and $C = 1\ \mu\mu\text{F}$, the rise time is $t_r = 2.2$ nsec, and the output waveform is readily sketched as shown in Fig. 15-7*b*. However, if $C = 10\ \mu\mu\text{F}$, then the rise time is $t_r = 22$ nsec, and the corresponding output waveform is as shown in Fig. 15-7*c*. Comparing the waveforms shown in Fig. 15-7, it is clear that to minimize distortion, the rise time must be as small as possible. Since the parasitic capacitance is always present in a physical circuit, it is impossible to make t_r zero, and hence it is impossible to obtain a perfect square wave.

3. Find the current $i_o(t)$ of the parallel R-L network shown in Fig. 15-8*a* with the periodic stepwise signal $i_s(t)$ shown in Fig. 15-8*b*.

From KCL we obtain

$$i_o(t) = i_s(t) - i_L(t) \tag{15-9}$$

Fig. 15-7. An example illustrating the significance of the rise time.



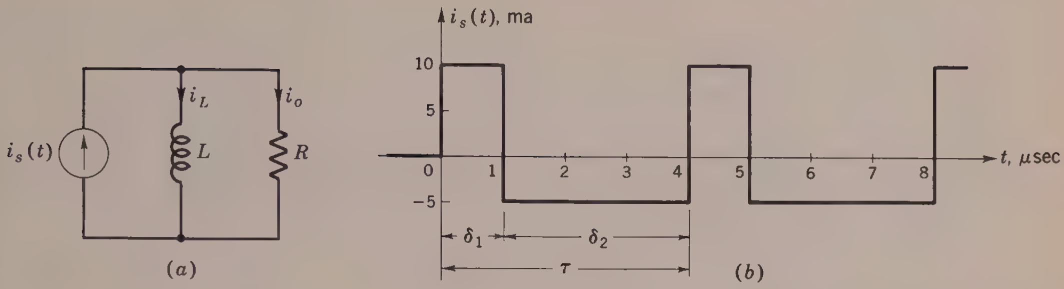


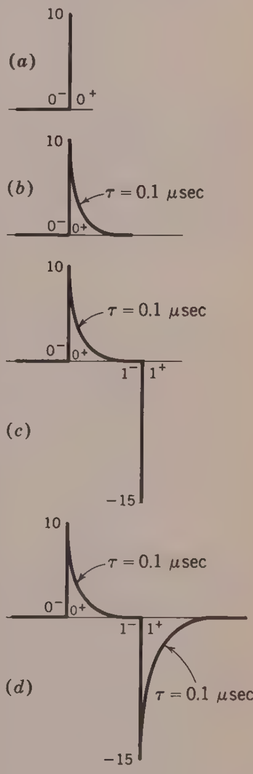
Fig. 15-8. A linear switching network driven by a periodic stepwise signal.

Since $i_s(t)$ jumps abruptly twice every cycle, and since the inductor current $i_L(t)$ cannot change instantaneously, Eq. (15-9) requires that any abrupt change in $i_s(t)$ must be absorbed by $i_o(t)$. Therefore, whenever $i_s(t)$ jumps by ΔI, $i_o(t)$ must also jump by ΔI. In order to illustrate the effect of the time constant on the waveform for $i_o(t)$, we shall consider two numerical cases.

Case 1: Small time constant Let $R = 10$ kΩ and $L = 1$ mH. The time constant is given by $\tau = L/R = 0.1$ μsec. Since $i_s(t) = 0$ for $t < 0$, the network is initially in equilibrium, and the inductor can be replaced by a short circuit. Therefore $i_o(0^-) = 0$. To find the transition state at $t = 0^+$, we observe that $i_o(t)$ must jump abruptly by an amount equal to the jump in $i_s(t)$; therefore, $i_o(0^+) = 10$ (Fig. 15-9*a*). The equilibrium state $i_o(t_e)$ for this network is always equal to zero because the current $i_s(t)$ will flow through the short-circuited inductor at equilibrium; hence $i_o(t_e) = 0$. The resulting exponential is shown in Fig. 15-9*b*. Observe that since the time constant here is much smaller than δ_1 of $i_s(t)$, the exponential waveform can be assumed to have reached steady state so that $i_o(1^-) = 0$. The transition state is, therefore, given by $i_o(1^+) = -15$ (Fig. 15-9*c*). The time constant remains unchanged, and the exponential waveform is shown in Fig. 15-9*d*. Again, since $\tau \ll \delta_2$, we can assume that the network has reached its equilibrium state, and so for $t > 4$, the waveform of $i_o(t)$ becomes periodic as shown in Fig. 15-10*a*. This solution is valid only when the time constant of the network is much smaller than the duration δ_1 and δ_2 of the stepwise signal.

Case 2: Large time constant Let $R = 1$ kΩ and $L = 10$ mH. The time constant becomes $\tau = L/R = 10$ μsec. As before, we

Fig. 15-9. The step-by-step procedure for analyzing a linear switching network consists of finding the transition state, the equilibrium state, and the time constant.



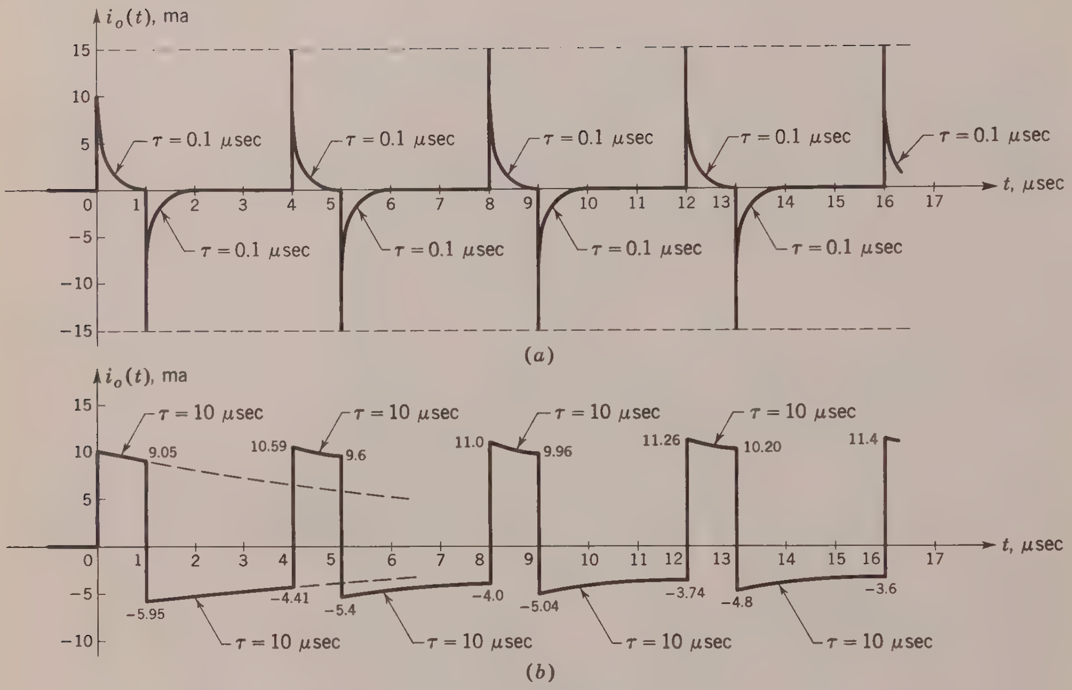


Fig. 15-10. The solution corresponding to a periodic input shown (*a*) for the case when the time constant is small and (*b*) for the case when the time constant is large.

have $i_o(0^-) = 0$, $i_o(0^+) = 10$, and $i_o(t_e) = 0$. Therefore, the solution is given by

$$i_o(t) = 10e^{-(t/10)} \qquad 0 \le t < 1 \text{ μsec} \tag{15-10}$$

and is plotted in Fig. 15-10*b*. Since the time constant in this case is large compared with δ_1, $i_o(t)$ will not reach equilibrium at $t = 1$. In fact, from Eq. (15-10) we obtain $i_o(1^-) = 9.05$. Since $I_s(t)$ jumps from 10 to -5 at $t = 1$ μsec, $\Delta I = -15$, and the transition state at $t = 1^+$ μsec is

$$i_o(1^+) = 9.05 - 15 = -5.95$$

The solution for $t \ge 1$ is given by

$$i_o(t) = -5.95e^{-[(t-1)/10]} \qquad 1 \le t < 4 \text{ μsec} \tag{15-11}$$

This waveform is shown in Fig. 15-10*b*. Observe again that since the time constant is large compared with δ_2, $i_o(t)$ will not reach equilibrium when the input signal jumps back to $i_s(t) = 10$ at $t = 4$ μsec. Instead, we have $i_o(4^-) = -4.41$, and hence

$$i_o(4^+) = 15 - 4.41 = 10.59$$

The solution is given by

$$i_o(t) = 10.59e^{-[(t-4)/10]} \qquad 4 \le t < 5 \ \mu\text{sec} \tag{15-12}$$

and is shown in Fig. 15-10*b*.

If we continue the above procedure over a few periods of the input signal, we obtain

$$i_o(5^-) = 9.6 \qquad i_o(5^+) = -5.4 \qquad i_o(t) = -5.4e^{-[(t-5)/10]} \qquad 5 \le t < 8 \ \mu\text{sec}$$

$$i_o(8^-) = -4 \qquad i_o(8^+) = 11 \qquad i_o(t) = 11e^{-[(t-8)/10]} \qquad 8 \le t < 9 \ \mu\text{sec}$$

$$i_o(9^-) = 9.96 \qquad i_o(9^+) = -5.04 \qquad i_o(t) = -5.04e^{-[(t-9)/10]} \qquad 9 \le t < 12 \ \mu\text{sec}$$

$$i_o(12^-) = -3.74 \qquad i_o(12^+) = 11.26 \qquad i_o(t) = 11.26e^{-[(t-12)/10]} \qquad 12 \le t < 13 \ \mu\text{sec}$$

. . . .

This waveform is shown in Fig. 15-10*b* over the interval $0 \le t < 12$. Observe that while the waveform shown in Fig. 15-10*a* becomes periodic after $t = 1$ μsec, the present waveform will become periodic only after a sufficiently long time interval.

Exercise 1: Under the same assumptions as in Example 1, find the voltage v_4 across the resistor R_4 in Fig. 15-3*a* by two methods: (*a*) Using the exponential waveform solution theorem. (*b*) Using the solution waveforms $v_C(t)$ and $v_o(t)$ shown in Fig. 15-3*e* and *f*.

Exercise 2: Using the exponential waveform solution theorem, find the current i_2 in the resistor R_2 in Fig. 15-3*a*. Check your solution by making use of the waveforms for $v_C(t)$ and $v_o(t)$ shown in Fig. 15-3*e* and *f*.

Exercise 3: Find the current $i_C(t)$ in Fig. 15-3*a* without finding the voltage $v_C(t)$. Check your solution using $v_C(t)$ in Fig. 15-3*e*.

Exercise 4: Using the exponential waveform solution theorem, find the voltage between the negative terminal of the capacitor and the positive terminal of the battery. Check your solution using $v_o(t)$ in Fig. 15-3*f*.

Exercise 5: Replace the capacitor in Fig. 15-5*a* by an L-henry linear inductor and find the voltage and current waveforms associated with this inductor for the small and large time-constant cases.

Exercise 6: Replace the inductor in Fig. 15-8*a* by a C-farad linear capacitor and find the voltage and current waveforms associated with this capacitor for the small and large time-constant cases.

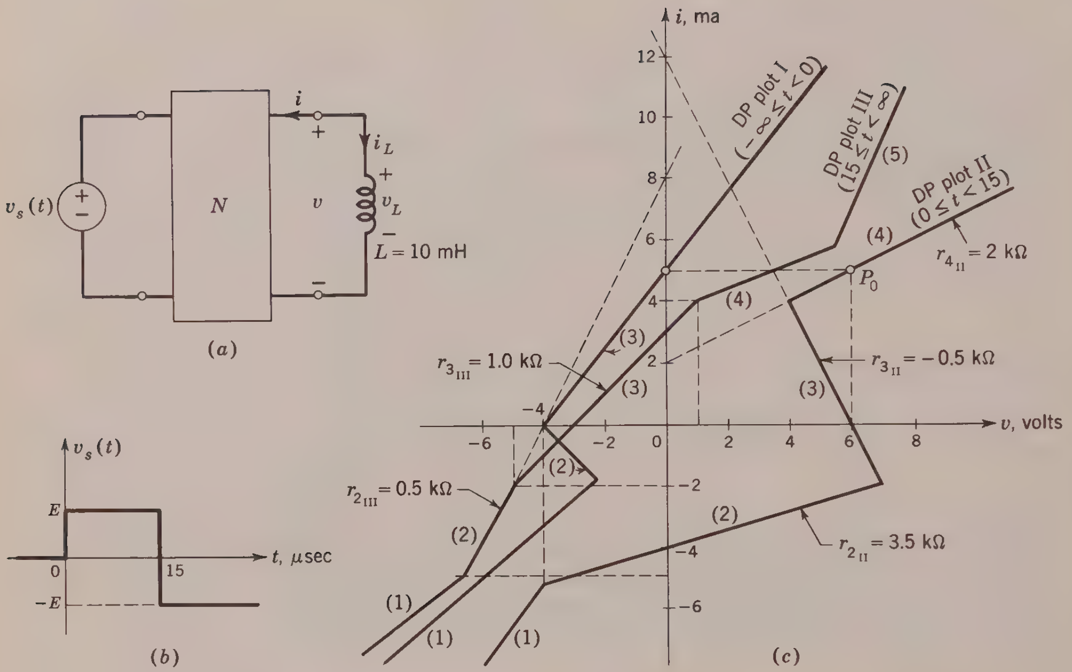


Fig. 15-11. A typical first-order nonlinear switching network.

15-3 ANALYSIS OF FIRST-ORDER NONLINEAR SWITCHING NETWORKS

Any first-order nonlinear switching network can be analyzed as a sequence of linear switching networks. The switching from one linear network to another can be handled by the principles presented in Chap. 14. Although nothing essentially new is involved, it is instructive to consider a numerical example in detail.

EXAMPLE

Consider the nonlinear switching network shown in Fig. 15-11*a* and the switching signal shown in Fig. 15-11*b*. The DP plots seen by the inductor during the three time intervals $-\infty \leq t < 0$, $0 \leq t < 15$, and $15 \leq t < \infty$ are shown in Fig. 15-11*c*. As usual, the basic problem is to find $i(t)$ and $v(t)$ across the inductor.

Since $v_s(t) = 0$ for all times $t < 0$, the network can be assumed to be in equilibrium. In particular, the equilibrium current i at $t = 0^-$ can be found from the intersection of DP plot I with the $v = 0$ axis; namely, $i(0^-) = 5$ ma. At $t = 0^+$, DP plot I switches to DP plot II. Since the inductor current cannot change instantaneously, $i(0^+) = i(0^-) = 5$ ma and the network must be operating at point P_0 on DP plot II. Our first task is to draw the dynamic route starting from P_0 at $t = 0^+$. However, we cannot draw the complete dynamic route for all times $t \geq 0$ because at $t = 15$ μsec we switch from DP plot II to DP plot III, and we do not yet know exactly at what point to switch from one to the other. Hence, it is necessary to determine $i(t)$ along DP plot II before we can complete the dynamic route.

The dynamic route along DP plot II is shown in Fig. 15-12*a*. Applying the principles in Chap. 14, we readily obtain the solution $i(t)$ as shown in Fig. 15-12*b*. The corresponding equations can be written by inspection, namely, $i(t)$ as shown in Fig. 15-12*b*.

$$i(t) = \begin{cases} 2 + 3e^{-(t/5)} & 0 \leq t < t_a \\ 12 - 8e^{+[(t-t_a)/20]} & t_a \leq t < t_b \\ -4 + 2e^{-[(t-t_b)/2.85]} & t_b \leq t < \infty \end{cases} \tag{15-13}$$

where

$$t_a = 5 \ln \frac{5-2}{4-2} = 2.03 \ \mu\text{sec}$$

and

$$t_b = 2.03 + 20 \ln \frac{-2-12}{4-12} = 13.23 \ \mu\text{sec}$$

Since $t_b < 15$ μsec, and since it will take an infinite time interval for $i(t)$ to attain the equilibrium value, we conclude that the

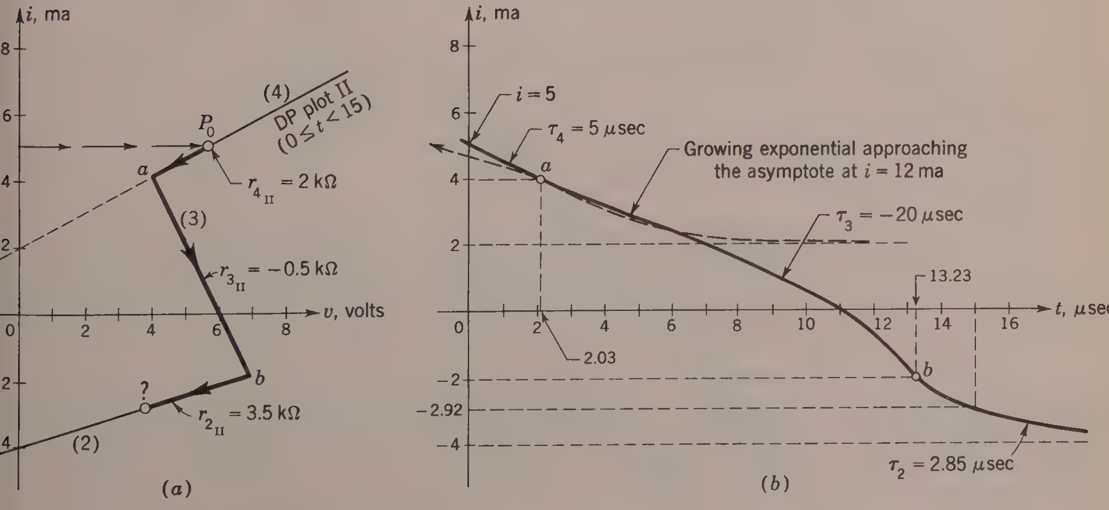


Fig. 15-12. The current waveform is obtained by inspection of the dynamic route along DP plot II.

dynamic route must terminate at some intermediate point on segment 2 as shown in Fig. 15-12*a*. This point can be found by substituting $t = 15\ \mu\text{sec}$ into Eq. (15-13).

$$i(15) = -4 + 2e^{-[(15-13.23)/2.85]} = -2.92 \text{ ma} \tag{15-14}$$

This terminal state is located at point c on DP plot II as shown in Fig. 15-13.

As soon as the solution reaches point c, DP plot II switches to DP plot III as shown in Fig. 15-13. Since the inductor current cannot change instantaneously, the transition state will be at point d of DP plot III. The remaining portion of the dynamic route can now be completed as shown in Fig. 15-13. Again, applying the principles of Chap. 14, we obtain the complete solution for $i(t)$ as shown in Fig. 15-14*a*. The corresponding solution $v_L(t)$ can now be easily determined either from $v_L = -L\,di/dt$ or directly from the dynamic route. This is shown in Fig. 15-14*b*.

Exercise 1: Find the voltage $v_L(t)$ and current $i(t)$ of the network in Fig. 15-11*a* corresponding to an initial inductor current $i_L(0) = 4$ ma.

Exercise 2: Replace the inductor in Fig. 15-11*a* by a 1,000-$\mu\mu$F linear capacitor and find the current and voltage waveforms associated with this capacitor for $t \geq 0$.

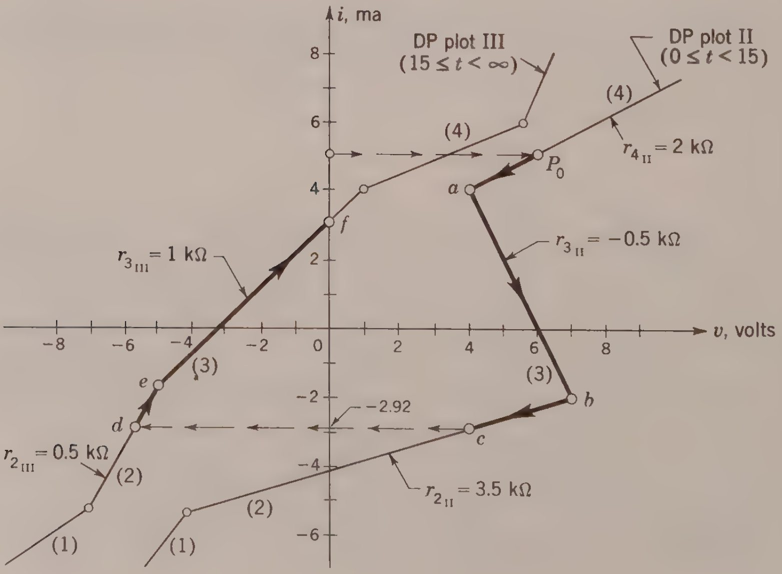


Fig. 15-13. The complete dynamic route for the nonlinear switching network in Fig. 15-11.

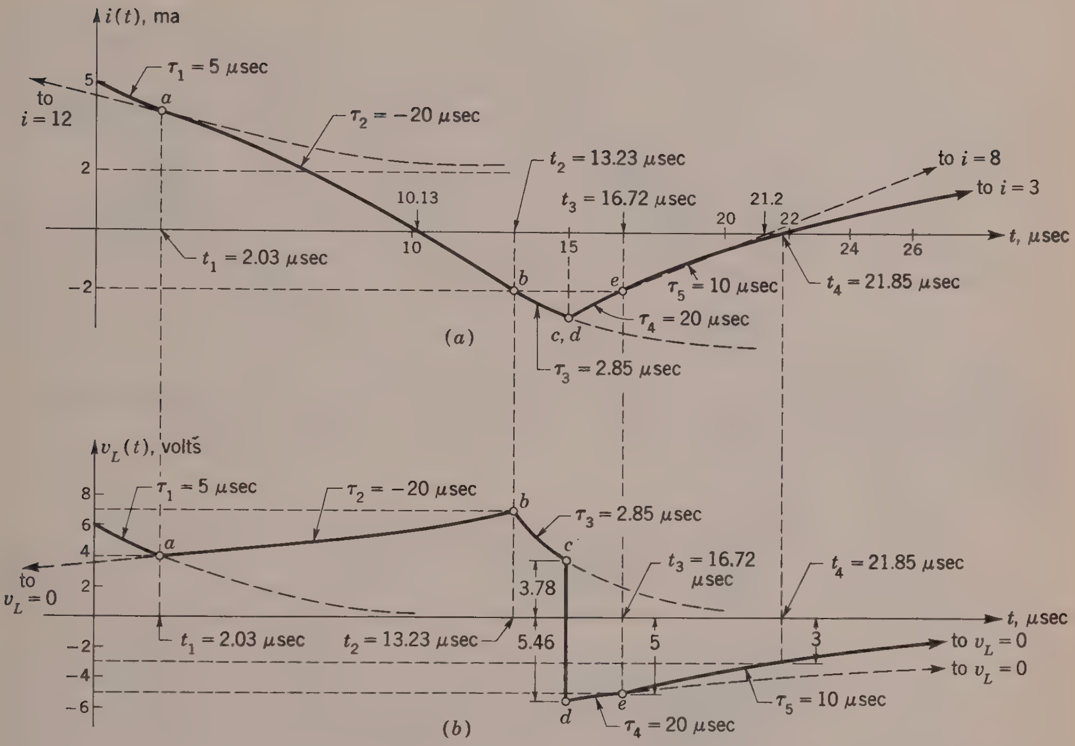


Fig. 15-14. The complete solution waveforms of $i(t)$ and $v_L(t)$ for the nonlinear switching network in Fig. 15-11.

15-4 ANALYSIS OF AUTONOMOUS FIRST-ORDER NETWORKS CONTAINING ONE NONLINEAR INDUCTOR OR CAPACITOR

We have already shown in Chap. 14 that an autonomous first-order network containing a *nonlinear* energy-storage element can be analyzed as a nonlinear switching network. We shall now illustrate the detailed procedure with an example.

EXAMPLE

Consider the autonomous nonlinear network shown in Fig. 15-15*a*. The DP plot of the resistive subnetwork *N* is shown in Fig. 15-15*b*, and the i_L-φ_L curve of the nonlinear inductor is shown in Fig. 15-15*c*. If $i_L(0) = 6$ ma, find the current $i(t)$ and the voltage $v(t)$ across the inductor for all times $t \geq 0$.

As usual, the first step is to draw the dynamic route. Since there are now two piecewise-linear characteristics, it is necessary to draw one dynamic route along each curve. The initial state at $t = 0$ corresponds to the point $P_0(i = -6$ ma) on the DP plot and

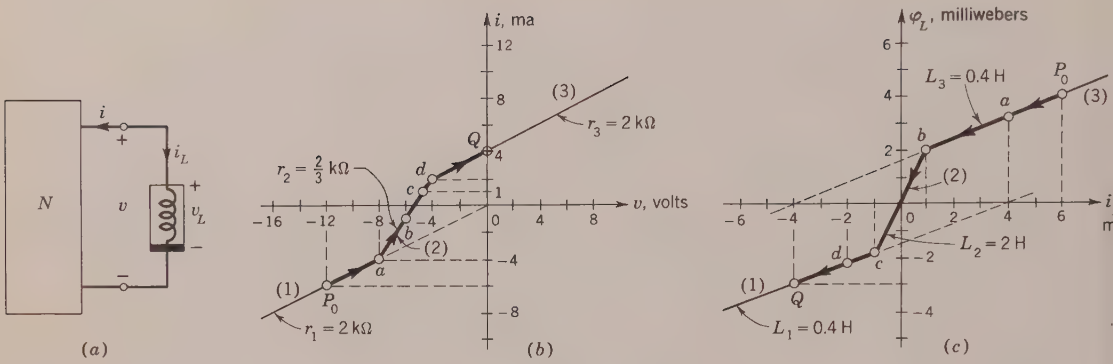


Fig. 15-15. An autonomous first-order network containing both a nonlinear resistive subnetwork and a nonlinear inductor.

the point $P_0(i_L = 6$ ma) on the i_L-φ_L curve. The dynamic route for $t \geq 0$ can be drawn with the help of the relationship

$$\frac{d\varphi_L}{dt} = v_L = v \qquad \begin{matrix} >0 & \text{whenever } v > 0 \\ <0 & \text{whenever } v < 0 \end{matrix} \tag{15-15}$$

From Eq. (15-15) we conclude that the dynamic route along the i_L-φ_L curve must be in the decreasing φ_L direction whenever the operating point along the v-i curve is in the left half-plane. Applying this simple rule, we immediately obtain the dynamic routes shown in Fig. 15-15*b* and *c*. Observe that since $i = -i_L$, the points *a*, *b*, *c*, *d*, and *Q* along the dynamic route of Fig. 15-15*b* must

TABLE 15-1 *Summary of information pertaining to the dynamic route in Fig. 15-15.*

Time interval, msec	Pertinent segment of v-i curve	Pertinent segment of i_L-φ_L curve	Time constant τ_{jk}, msec	Initial and transition state, ma
$0 \leq t < t_a$	1	3	$\tau_{13} = \frac{0.4}{(2)(10^3)} = 0.2$	$i_L(0^-) = i_L(0^+) = 6$
$t_a \leq t < t_b$	2	3	$\tau_{23} = \frac{0.4}{(2/3)(10^3)} = 0.6$	$i_L(t_a^-) = i_L(t_a^+) = 4$
$t_b \leq t < t_c$	2	2	$\tau_{22} = \frac{2}{(2/3)(10^3)} = 3$	$i_L(t_b^-) = i_L(t_b^+) = 1$
$t_c \leq t < t_d$	2	1	$\tau_{21} = \frac{0.4}{(2/3)(10^3)} = 0.6$	$i_L(t_c^-) = i_L(t_c^+) = -1$
$t_d \leq t < t_Q$	3	1	$\tau_{31} = \frac{0.4}{(2)(10^3)} = 0.2$	$i_L(t_d^-) = i_L(t_d^+) = -2$

correspond to the points a, b, c, d, and Q along the dynamic route of Fig. 15-15c. If we denote the time when the solution reaches points a, b, c, d, and Q by t_a, t_b, t_c, t_d, and t_Q, respectively, then we can summarize the pertinent data as shown in Table 15-1.

Observe that for $t_a \leq t < t_b$, the inductor is operating on segment 3. Hence, the network can be replaced by the equivalent circuit shown in Fig. 15-16a. Similarly, corresponding to the time intervals $(t_b \leq t < t_d)$ and $(t_d \leq t < t_Q)$, we have the equivalent circuits shown in Fig. 15-16b and c, respectively. The composite networks N_1, N_2, and N_3 are characterized by the v_1-i_1, v_2-i_2, and v_3-i_3 curves shown in Fig. 15-16. It is important to observe that the currents i_1, i_2, i_3 are *not* equal to the actual current i_L in the nonlinear inductor but are related by KCL.

$$i_j(t) = I_j - i_L(t) \qquad j = 1, 2, 3 \tag{15-16}$$

Therefore, to translate the dynamic route along the i_L-φ_L curve of Fig. 15-15c into the corresponding dynamic routes along the v_1-i_1, v_2-i_2, and v_3-i_3 curves, it is necessary to determine the equivalent initial and terminal states.

Let us start with the v_1-i_1 curve of Fig. 15-16a. This DP plot is valid during the time interval $0 \leq t < t_b$. From the dynamic route along the i_L-φ_L curve of Fig. 15-15c, we obtain the initial state $i_L(0^+) = 6$ ma and the terminal state $i_L(t_b^-) = 1$ ma. If we substitute these values into Eq. (15-16), we obtain

$$i_1(0^+) = I_1 - i_L(0^+) = -4 - 6 = -10 \text{ ma} \tag{15-17}$$

and

$$i_1(t_b^-) = I_1 - i_L(t_b^-) = -4 - 1 = -5 \text{ ma} \tag{15-18}$$

The corresponding dynamic route along the v_1-i_1 curve must, therefore, start at point $P_0(i_1 = -10)$ and terminate at point $b(i_1 = -5)$ as shown by the solid paths in Fig. 15-16a.

At $t = t_b^+$ we switch from the v_1-i_1 curve to the v_2-i_2 curve shown in Fig. 15-16b. This DP plot is valid during the time interval $t_b \leq t < t_c$. From the dynamic route along the i_L-φ_L curve of Fig. 15-15c we obtain the initial state $i_L(t_b^+) = i_L(t_b^-) = 1$ ma and the terminal state $i_L(t_c^-) = -1$ ma. If we substitute these values into Eq. (15-16), we obtain

$$i_2(t_b^+) = I_2 - i_L(t_b^+) = 0 - (1) = -1 \text{ ma} \tag{15-19}$$

$$i_2(t_c^-) = I_2 - i_L(t_c^-) = 0 - (-1) = 1 \text{ ma} \tag{15-20}$$

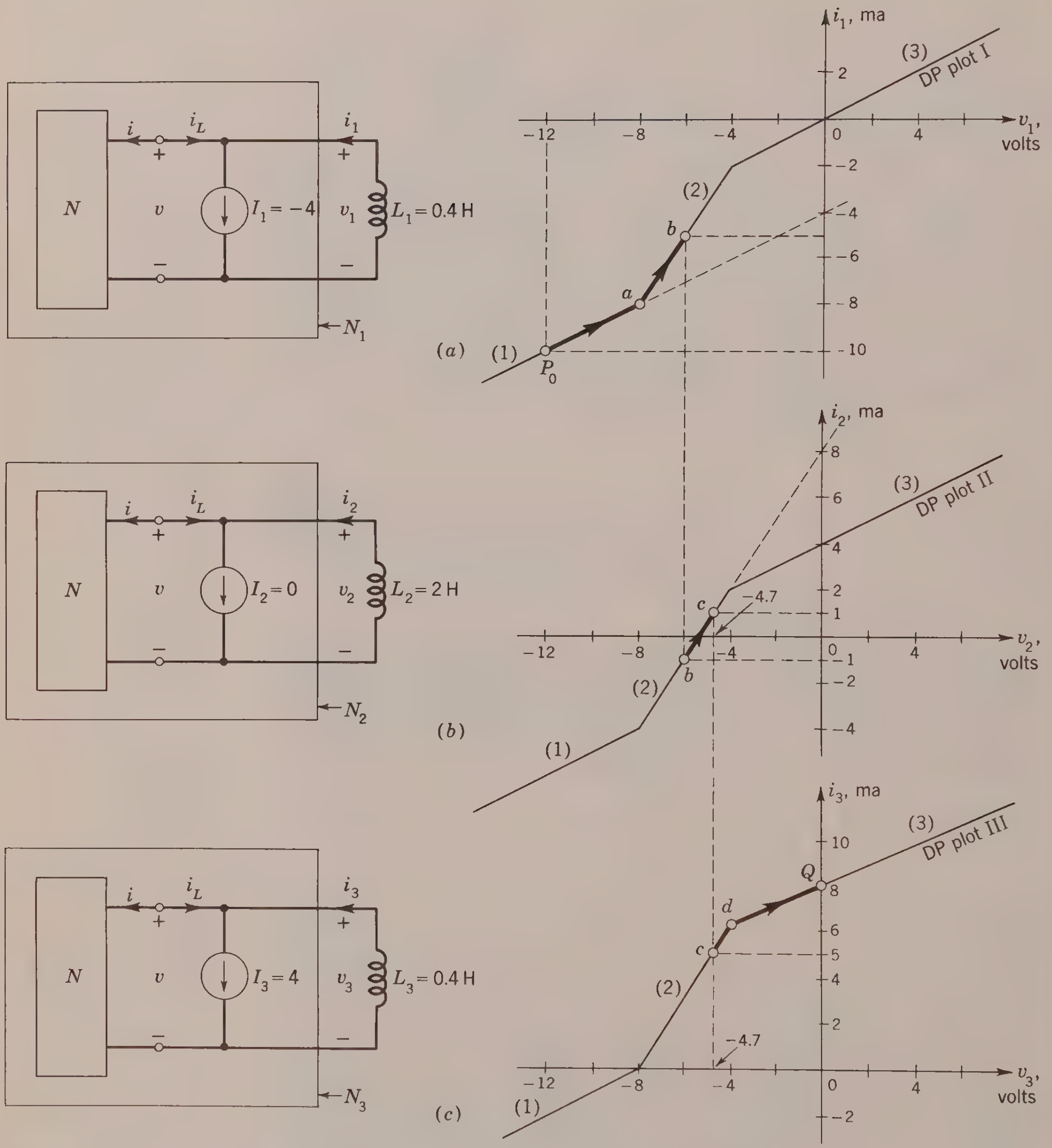


Fig. 15-16. The three equivalent circuits for the network shown in Fig. 15-15 corresponding to the time intervals $(0, t_b)$, (t_b, t_c), and (t_c, ∞).

The corresponding dynamic route along the v_2-i_2 curve must, therefore, start at point $b(i_2 = -1)$ and terminate at point $c(i_2 = 1)$ as shown by the solid paths in Fig. 15-16*b*.

At $t = t_c^+$ we switch from the v_2-i_2 curve to the v_3-i_3 curve shown in Fig. 15-16*c*. This DP plot is valid during the time intervals

$t_c \leq t < \infty$. From the dynamic route along the i_L-φ_L curve of Fig. 15-15c we obtain the initial state $i_L(t_c^+) = i_L(t_c^-) = -1$ ma and the equilibrium state $i_L(t_e) = -4$ ma. If we substitute these values into Eq. (15-16), we obtain

$$i_3(t_c^+) = I_3 - i_L(t_c^+) = 4 - (-1) = 5 \text{ ma} \tag{15-21}$$

$$i_3(\infty) = I_3 - i_L(\infty) = 4 - (-4) = 8 \text{ ma} \tag{15-22}$$

The corresponding dynamic route along the v_3-i_3 curve must, therefore, start at point $c(i_3 = 5$ ma) and terminate at point $Q(i_3 = 8$ ma) as shown by the solid paths in Fig. 15-16c.

It is now a simple matter to solve for the waveforms of $i(t)$ and $v(t)$. This involves the determination of $i_1(t)$, $i_2(t)$, and $i_3(t)$ corresponding to the dynamic routes shown in Fig. 15-16a to c. Applying the principles of Chap. 14, we obtain the waveforms for $i_1(t)$, $i_2(t)$, and $i_3(t)$ as shown in Fig. 15-17a to c, respectively. The values of t_a, t_b, t_c, t_d, and t_Q are calculated from the data in Table 15-1 and the dynamic routes in Fig. 15-16.

$$t_a = 0 + \tau_{13} \ln \frac{i_1(0) - i_1(t_e)}{i_1(t_a) - i_1(t_e)} = 0.2 \ln \frac{-10 - (-4)}{-8 - (-4)} = 0.08 \text{ msec}$$

$$i_b = t_a + \tau_{23} \ln \frac{i_1(t_a) - i_1(t_e)}{i_1(t_b) - i_1(t_e)} = 0.08 + 0.6 \ln \frac{-8 - 4}{-5 - 4}$$
$$= 0.25 \text{ msec}$$

$$t_c = t_b + \tau_{22} \ln \frac{i_2(t_b) - i_2(t_e)}{i_2(t_c) - i_2(t_e)} = 0.25 + 3 \ln \frac{-1 - 8}{1 - 8} = 1.0 \text{ msec}$$

$$t_d = t_c + \tau_{21} \ln \frac{i_3(t_c) - i_3(t_e)}{i_3(t_d) - i_3(t_e)} = 1.0 + 0.6 \ln \frac{5 - 12}{6 - 12} = 1.1 \text{ msec}$$

$$t_Q = \infty$$

The final solution $i(t)$ can now be obtained from Eq. (15-16); namely,

$$i(t) = -i_L(t) = i_j(t) - I_j$$

$$i(t) = \begin{cases} i_1(t) - (-4) & 0 \leq t < t_b \\ i_2(t) - (0) & t_b \leq t < t_c \\ i_3(t) - 4 & t_c \leq t < t_Q \end{cases} \tag{15-23}$$

This is plotted in Fig. 15-18a. By a similar procedure we obtain the complete solution $v(t)$ as shown in Fig. 15-18b.

A closer examination of the waveforms shown in Figs. 15-17 and 15-18 will reveal that the current source in Fig. 15-16*a*, *b*, and *c* does not affect the time constant of each segment of $i(t)$. Moreover, observe that the *same* constant I_j *added* to the v_j-i_j curves in Fig. 15-16 is subsequently *subtracted* from the corresponding waveforms shown in Fig. 15-17. Therefore, the final solution waveforms shown in Fig. 15-18 could have been obtained directly from the dynamic route shown in Fig. 15-15*b* and *c*, provided the

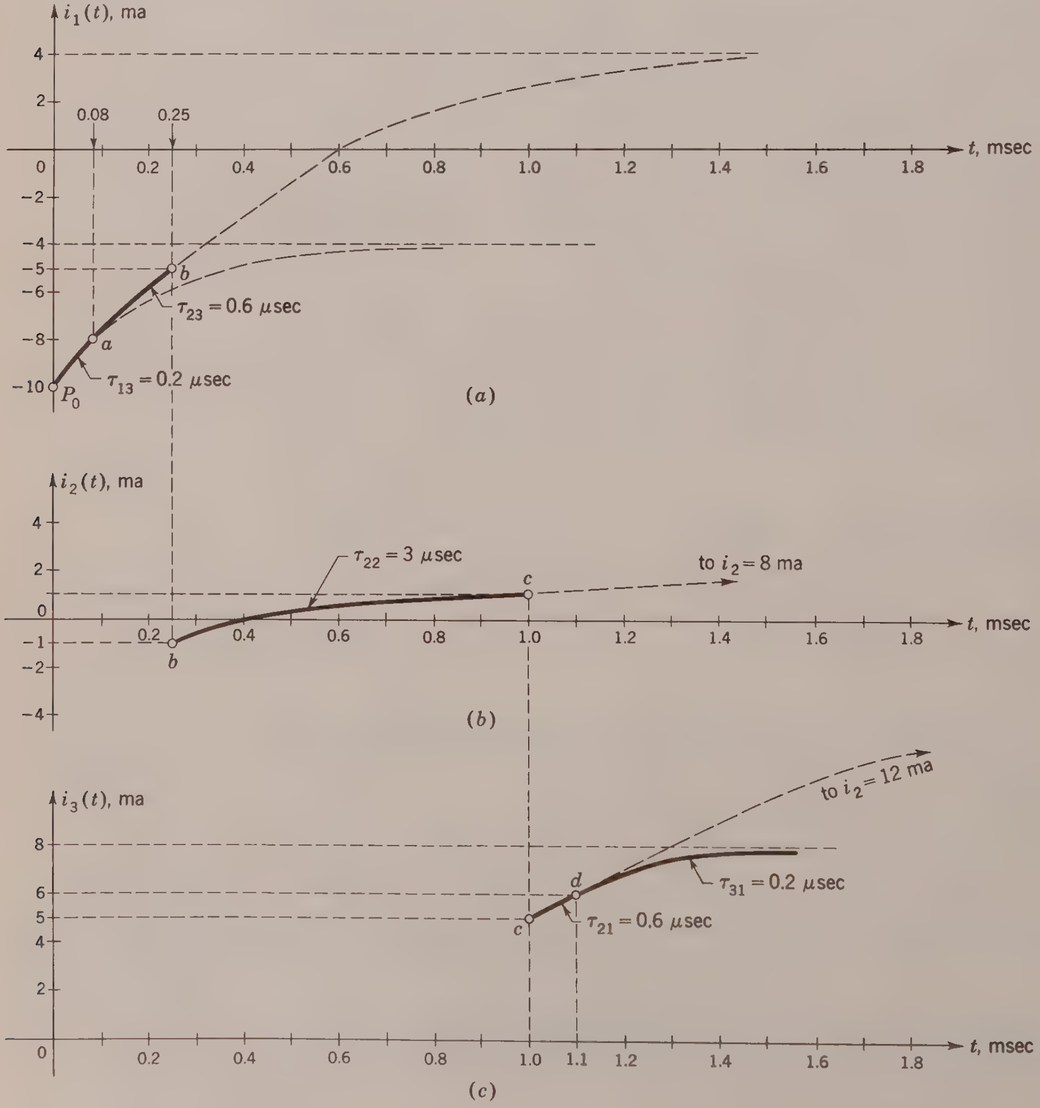


Fig. 15-17. The solution waveforms for $i_1(t)$, $i_2(t)$, and $i_3(t)$ over the time intervals (0,0.25), (0.25,1.0), and (1.0,∞).

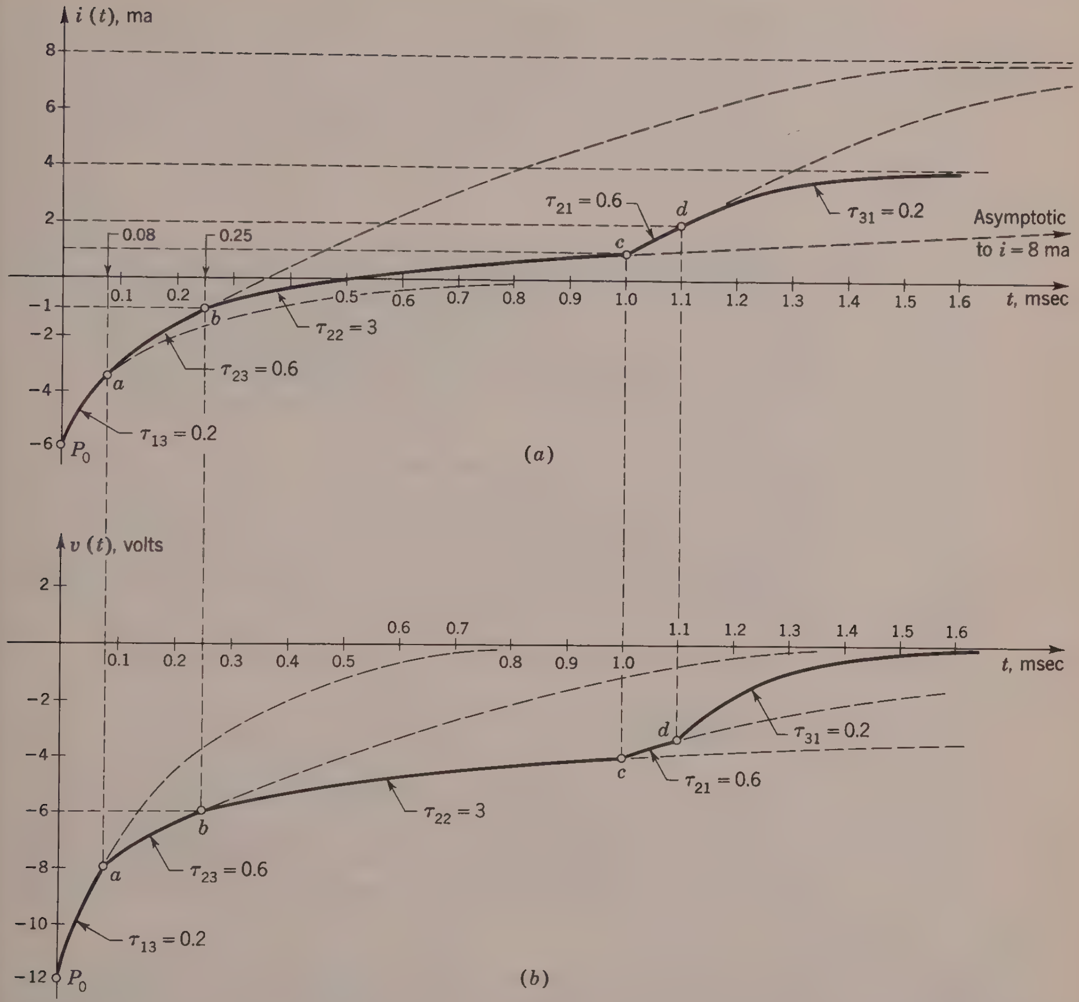


Fig. 15-18. The complete solution waveforms of $i(t)$ and $v(t)$ for the nonlinear network shown in Fig. 15-15.

appropriate time constant were chosen for each segment combination. In other words, the intermediate steps shown in Figs. 15-16 and 15-17 are really unnecessary. They are given only for pedagogical reasons.

There is an alternate procedure for solving the above example which is sometimes simpler to apply. This procedure chooses the flux linkage φ_L as the state variable. If we denote the v-i curve by $v = v(i)$ and the i_L-φ_L curve by $i_L = i_L(\varphi_L)$, then the normal-form equation is given by

$$\frac{d\varphi_L}{dt} = v(i) = v(-i_L(\varphi_L)) \tag{15-24}$$

or

$$\frac{d\varphi_L}{dt} = f(\varphi_L) \tag{15-25}$$

where $f(\varphi_L) \equiv v(-i_L(\varphi_L))$ can be obtained by a graphical composition between the v-i curve and the $(-i_L)$-φ_L curve. Clearly, the $f(\varphi_L)$ curve will also be piecewise-linear, and the solution $\varphi_L(t)$ can be obtained by the procedure developed in Chap. 14. Once $\varphi_L(t)$ is found, we can obtain $v(t) = d\varphi_L/dt$. The current $i(t)$ can be found by a graphical composition between the v-i curve and $v(t)$.

Exercise 1: Verify the solution waveforms $i(t)$ and $v(t)$ in Fig. 15-18 by the alternate procedure given by Eq. (15-25).

Exercise 2: Replace the inductor in Fig. 15-15*a* by a capacitor whose v-q curve is obtained by replacing the coordinates i_L and φ_L in Fig. 15-15*c* by v and q, respectively. Find the solution waveforms $i(t)$ and $v(t)$ across the capacitor by the method dual to that shown in Fig. 15-16.

Exercise 3: Repeat Exercise 2 by the alternate composition procedure.

15-5 SUMMARY

First-order switching networks A first-order network (linear or nonlinear) containing switches as well as stepwise signals is called a first-order switching network. This class of nonautonomous networks can be analyzed as a sequence of autonomous first-order networks.

The exponential waveform solution theorem The solution waveform of the voltage across any terminal pair or the current in any branch of a first-order linear switching network always consists of a sequence of exponential waveforms. The time constant of the exponential is the same for all voltage and current waveforms during the same time interval. This theorem allows us to find the voltage across any terminal pair or current in any branch of a first-order switching network by inspection. In other words, it is not necessary to find the capacitor voltage or the inductor current before finding the remaining variables.

A first-order autonomous network containing a nonlinear inductor or capacitor can be solved by the following two piecewise-linear methods:

1. As a first-order switching network containing a linear inductor or a linear capacitor
2. As a first-order autonomous network with the flux linkage or charge as the state variable

PROBLEMS

15-1 Consider the linear R-C network shown in Fig. P15-1a. The switching signal $v_s(t)$ is shown in Fig. P15-1b.

(a) Calculate the rise time of the network.

(b) Determine $v_o(t)$ for $t \leq 25$ μsec.

(c) Determine the steady-state waveform of $v_o(t)$.

(d) Repeat (c) with $C = 0.001$ μF.

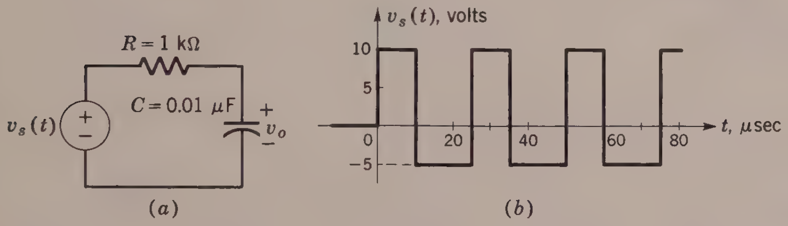


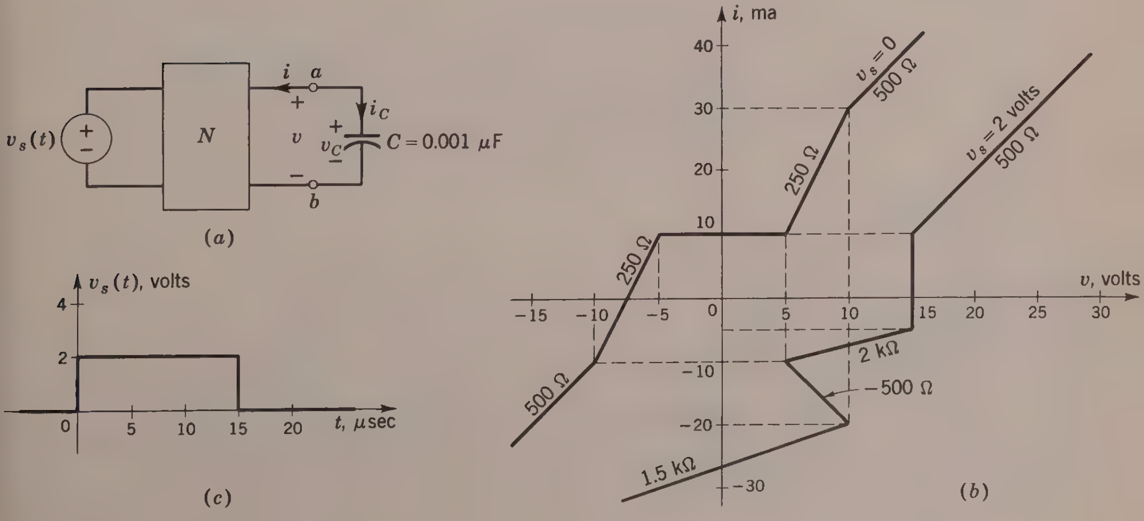
Fig. P15-1.

15-2 The DP plot of the switching network shown in Fig. P15-2a is given in Fig. P15-2b corresponding to the switching signal shown in Fig. P15-2c.

(a) Assuming that the network is at rest for $t \leq 0$, find $v_C(t)$ for $t \geq 0$.

(b) If the initial charge $q(0) = 2 \times 10^{-8}$ coul, find $i_C(t)$.

Fig. P15-2.



15-3 The DP plot of the switching network shown in Fig. P15-3*a* is given in Fig. P15-3*b* corresponding to the switching signal $v_s(t)$ shown in Fig. P15-3*c*. Find the waveform $v_L(t)$ across the inductor for $t \geq -2$ msec. The current $i_L(t)$ at $t = -2$ msec is known to be 5 ma.

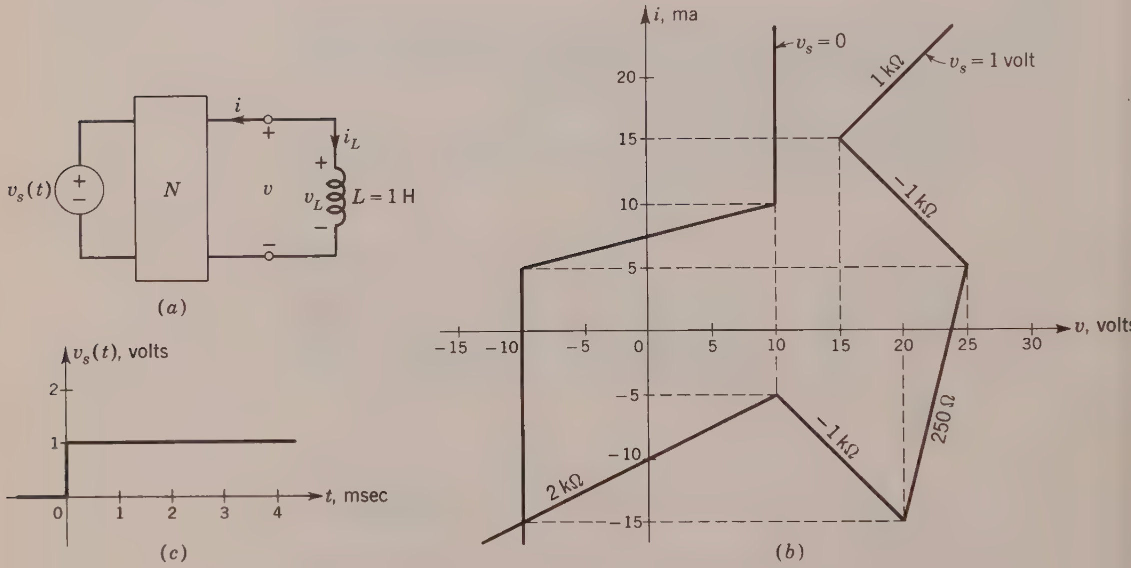


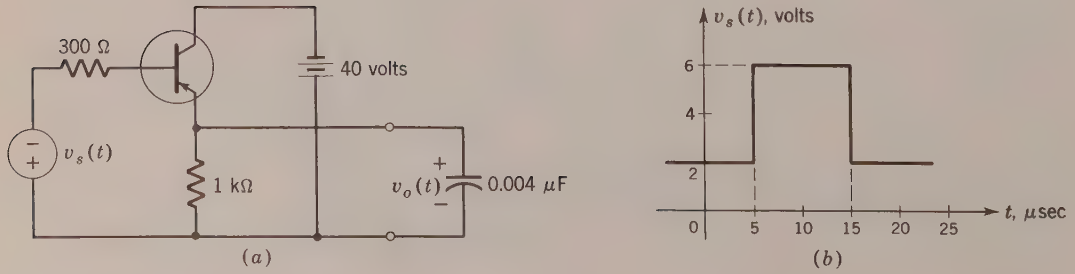
Fig. P15-3.

15-4 A first-order switching network and its switching signal $v_s(t)$ are shown in Fig. P15-4*a* and *b*.

(a) Using the *p-n-p* transistor model C_E with $\beta = 60$, $V_{BEO} = 0$, and $I_{CEO} = 0$, find the DP plot across the capacitor terminals during each switching time interval.

(b) Plot the waveform $v_o(t)$ for $t \geq 0$.

Fig. P15-4.



15-5 Consider the first-order switching circuit shown in Fig. P15-5*a* and the switching signal $v_s(t)$ shown in Fig. P15-5*b*. Let the FET be represented by a simplified version of an *n*-channel FET model A as shown in Fig. P15-5*c* with $k = -1.5$ ma per volt.

(a) Determine the DP plot across terminals *a-b* during each switching time interval.

(b) Plot the waveform $v_o(t)$ for $t \geq 0$.

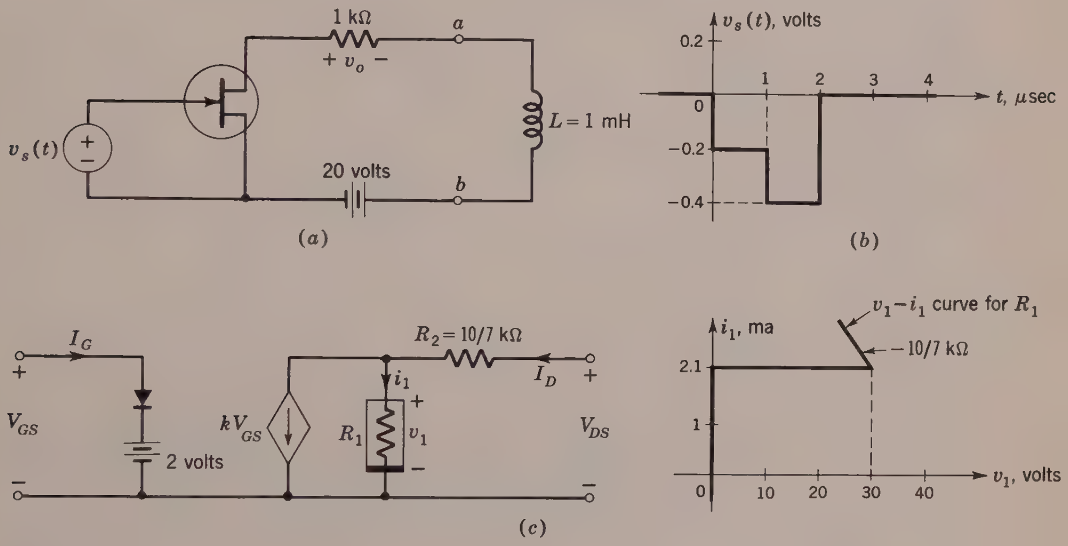


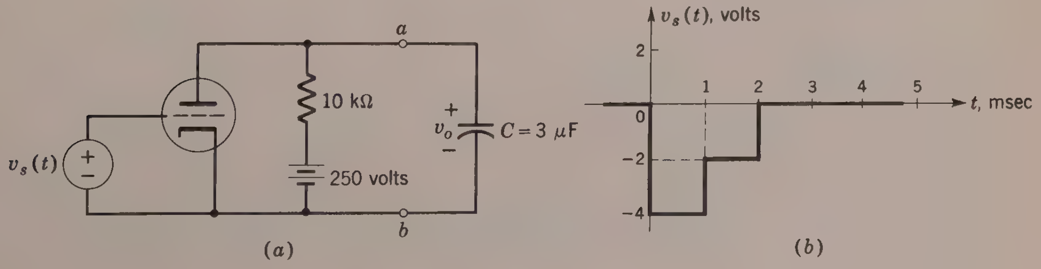
Fig. P15-5.

15-6 Consider the first-order switching network shown in Fig. P15-6*a* and the switching signal $v_s(t)$ shown in Fig. P15-6*b*.

(a) Using the triode model C with $r_g = 100\ \Omega$, $r_p = 2.0$ kΩ, and $\mu = 20$, find the DP plot across terminals *a-b*.

(b) Plot the output voltage $v_o(t)$ for $t \geq 0$.

Fig. P15-6.



15-7 A first-order switching network consisting of two identical triodes is shown in Fig. P15-7*a*. The associated switching signal $v_s(t)$ is shown in Fig. P15-7*b*.

(a) Using the triode model C with $r_g = 100\ \Omega$, $r_p = 10\ \text{k}\Omega$, and $\mu = 24$, find the DP plot across terminals *a-b*.

(b) Plot the voltage $v_L(t)$ for $t \geq 0$.

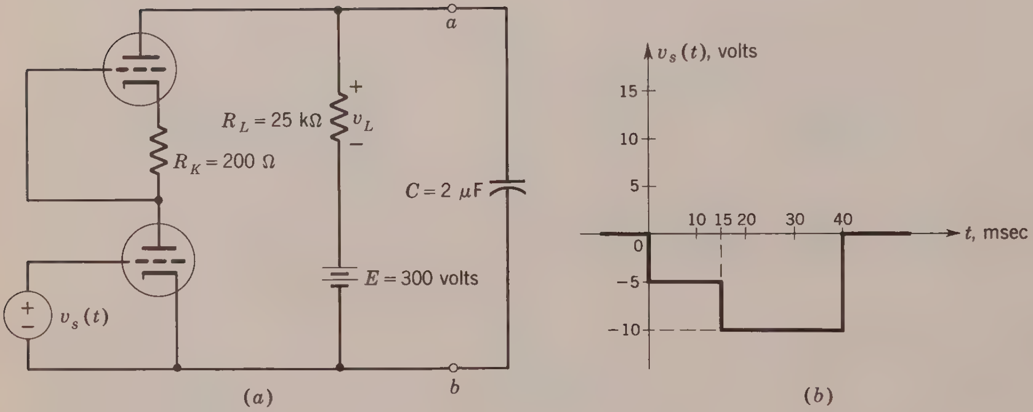


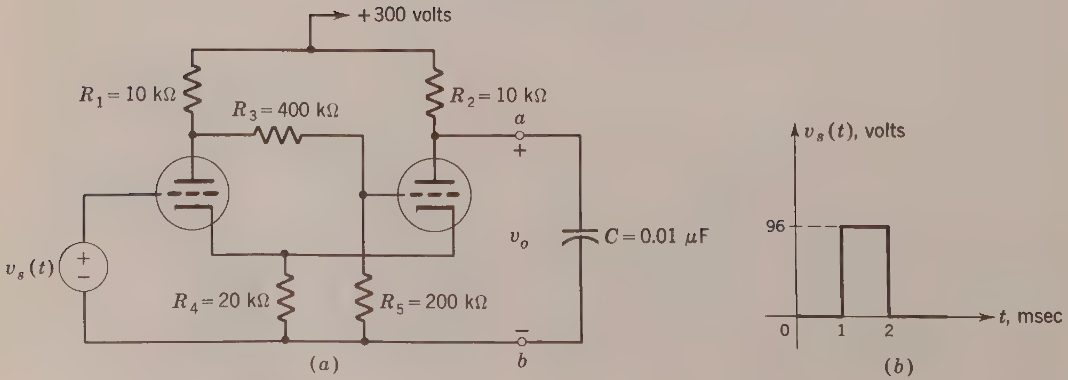
Fig. P15-7.

15-8 One of the most useful and versatile first-order switching networks is the Schmitt trigger circuit shown in Fig. P15-8*a*. A typical switching signal is shown in Fig. P15-8*b*.

(a) Assuming that the two identical triodes may be represented by the triode model C with $r_g = 200\ \Omega$, $r_p = 7.0\ \text{k}\Omega$, and $\mu = 24$, find the DP plot across terminals *a-b* during each switching time interval.

(b) Plot the output voltage waveform $v_o(t)$ for $t \geq 0$.

Fig. P15-8.


+300 volts
$R_1 = 10\ \text{k}\Omega$
$R_2 = 10\ \text{k}\Omega$
$R_3 = 400\ \text{k}\Omega$
a
$+$
v_o
$-$
b
$C = 0.01\ \mu\text{F}$
$v_s(t)$
$R_4 = 20\ \text{k}\Omega$
$R_5 = 200\ \text{k}\Omega$
(a)
$v_s(t)$, volts
96
0
1
2
t, msec
(b)

15-9 Consider the autonomous first-order network shown in Fig. P15-9*a*. The nonlinear resistive black box *N* is characterized by the DP plot shown in Fig. P15-9*b*, and the nonlinear inductor is characterized by the i_L-φ_L curve shown in Fig. P15-9*c*. Find the inductor current $i_L(t)$ and the inductor voltage $v_L(t)$ for $t \geq 0$ corresponding to an initial flux linkage $\varphi_L(0) = -6$.

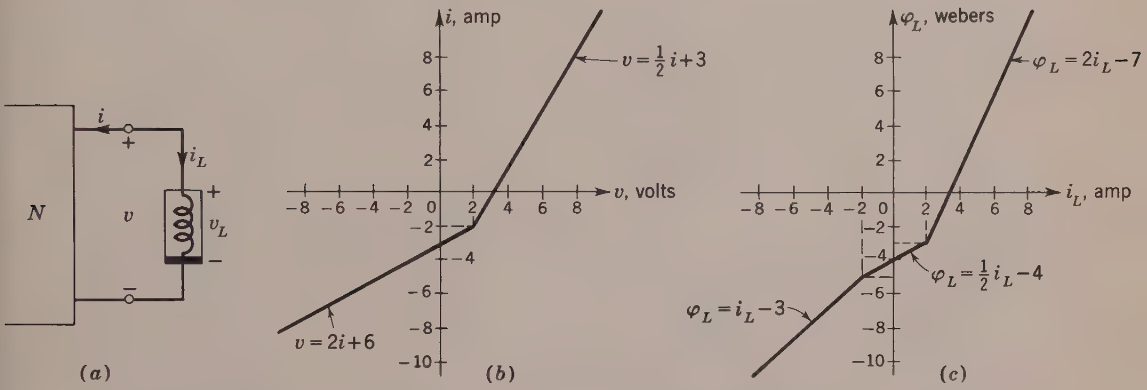


Fig. P15-9.

15-10 Consider the autonomous first-order network shown in Fig. P15-10*a*. The nonlinear resistive black box *N* is characterized by the DP plot shown in Fig. P15-10*b*, and the nonlinear capacitor is characterized by the *v*-*q* curve shown in Fig. P15-10*c*.

(a) Find the equilibrium states of this network by inspection.

(b) If the initial charge on the capacitor is given by $q(0) = -10$, find the voltage $v(t)$ and current $i(t)$ for $t \geq 0$.

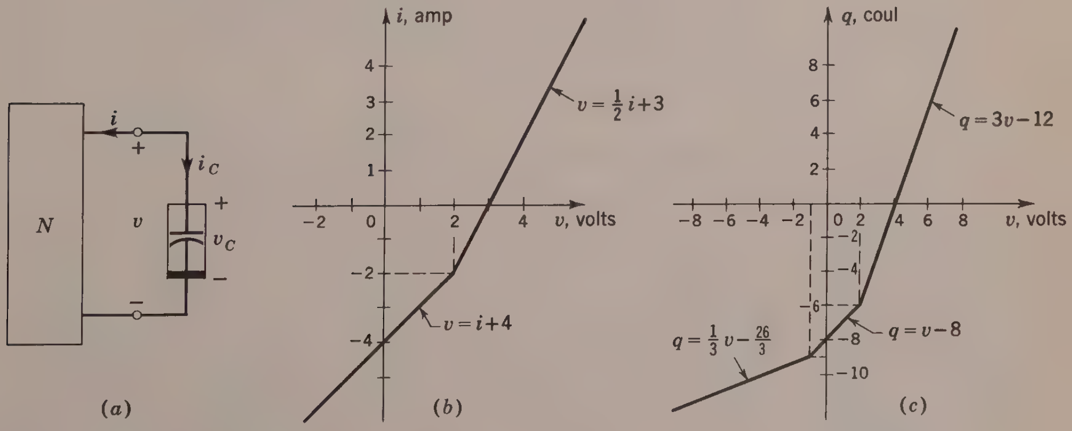


Fig. P15-10.

15-11 Consider the autonomous first-order network shown in Fig. P15-11*a*. The nonlinear resistive black box *N* is characterized by the DP plot shown in Fig. P15-11*b*, and the nonlinear inductor is characterized by the i_L-φ_L curve shown in Fig. P15-11*c*.

(a) Find the equilibrium states of this network by inspection.

(b) If the initial flux linkage is given by $\varphi_L(0) = 9$, find the current $i(t)$ and voltage $v(t)$ for $t \geq 0$.

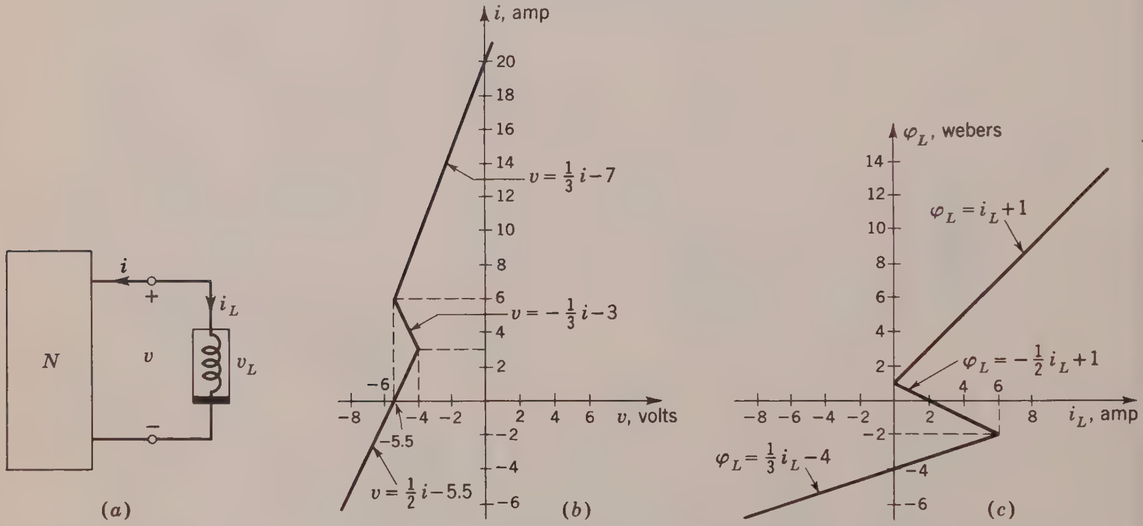


Fig. P15-11.

15-12 An important problem in the analysis of integrated circuits is to determine the voltage response of a nonlinear capacitor when it is connected to a stepwise signal in series with a linear resistor as shown in Fig. P15-12*a*. Let the capacitor be characterized by the *v*-*q* curve shown in Fig. P15-12*b*.

(a) Find $v_C(t)$ for $t \geq 0$ if the switching signal is given by

$$v_s(t) = \begin{cases} 20 & t \geq 0 \\ 0 & t < 0 \end{cases}$$

(b) Find $v_C(t)$ for $t \geq 0$ if the switching signal is given by

$$v_s(t) = \begin{cases} 0 & t \geq 0 \\ 20 & t < 0 \end{cases}$$

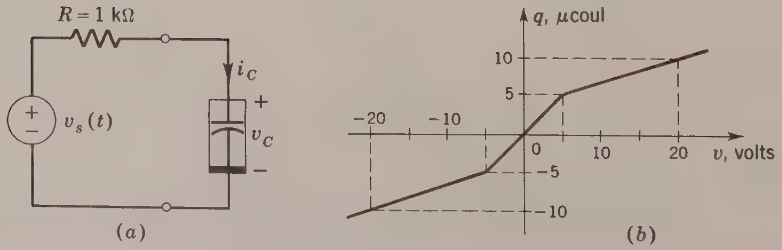


Fig. P15-12.

16 SYNTHESIS OF FIRST-ORDER MULTIVIBRATORS: THE BLACK-BOX APPROACH

16-1 WHAT IS A FIRST-ORDER MULTIVIBRATOR?

Multivibrator is a generic name normally given to an important class of dynamic nonlinear networks which are found in virtually any modern electronic system of moderate complexity. Multivibrators are indispensable building blocks of pulse and digital systems such as digital computers and radar. It is not our purpose here to undertake an exhaustive study of the countless variations of multivibrator circuits, for this would only lead to confusion, if not frustration. Instead, we shall study the general concepts underlying the operation of all *first-order multivibrators.* We shall be particularly interested in what network components can be used to synthesize a multivibrator and how to use element characteristics to meet a given set of specifications. Once a basic multivibrator circuit is synthesized, there are many "finishing touches" that we can apply to modify the circuit and improve its performance. The ability to apply the appropriate finishing touches will come naturally after the principles of operation and the roles played by the circuit elements are clearly understood.

For complete generality, we shall define a multivibrator as any dynamic network (of arbitrary order of complexity) having the property that at least one voltage or one current waveform associated with the network exhibits an *abrupt jump*[1] either periodically or whenever a relatively small switching signal known as the *triggering source* is applied. Since most practical multivibrators can be considered as first-order switching networks, we shall study only first-order multivibrators in this chapter. Depending on the type of the DP plot seen by the energy-storage element, the class of first-order multivibrators can be classified into three subclasses: *astable, monostable,* or *bistable*. The following three sections will be devoted to a detailed presentation of these three subclasses.

[1] As will be shown later, no *physical* network can exhibit an instantaneous change in either the voltage or the current waveforms associated with the network. In order to include practical multivibrators in our definition, we shall interpret the term *abrupt jump* to be a very fast change in the waveform over an extremely short interval of time.

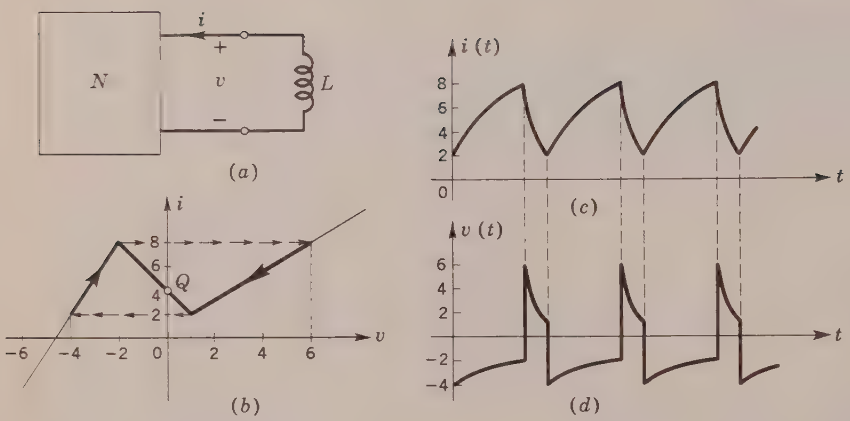


Fig. 16-1. The DP plot seen by the inductor of an astable multivibrator must be voltage-controlled, and the negative-resistance portions of the curve must intersect the *i* axis.

16-2 ASTABLE (FREE-RUNNING) MULTIVIBRATORS

An astable (free-running) multivibrator is a first-order autonomous nonlinear network with three distinguishing properties: (1) it has one, and only one, equilibrium state Q, (2) the equilibrium state Q is unstable, and (3) the DP plot of the resistive subnetwork is voltage-controlled if the energy-storage element is an inductor and current-controlled if the energy-storage element is a capacitor. For example, the circuits shown in Figs. 16-1 and 16-2 satisfy these properties and are, therefore, astable multivibrators. Observe that in both networks the slope of the DP plot at the equilibrium state Q is negative. Since this condition is necessary for Q to be unstable, we can conclude that in order to satisfy the properties defining an astable multivibrator, it is necessary that the DP plot exhibit a negative-resistance region. Moreover, the negative-resistance portion must intersect the i axis if the energy-storage element is an inductor and the v axis if the energy-storage element is a capacitor.

From the dynamic routes and waveforms shown in Figs. 16-1 and 16-2, we can conclude that both the current and voltage solutions of any astable multivibrator are periodic time functions. Moreover, the inductor voltage waveform and the capacitor current waveform exhibit two abrupt jumps per cycle.[1] Observe that over each half-cycle the dynamic route stays entirely on one branch of the DP plot. In earlier days, the resistive subnetwork N of astable multivibrators usually consisted of two vacuum tubes connected in such a way that during each half-cycle one tube was conducting while the other was not. Therefore, while one tube was

---

[1] In practice, the parasitic inductance in series with any physical capacitor will prevent an abrupt jump in the capacitor current waveform. To show this, let $v_L = L\, di_C/dt$ be the voltage drop across the parasitic inductor (with inductance L) due to the capacitor current $i_C(t)$ flowing through it. Now, an abrupt jump in $i_C(t)$ at any time $t = t_j$ means that $di_C/dt = \infty$ at $t = t_j$. This implies that $v_L(t_j) = \infty$, which is impossible in view of the principle of conservation of energy. By dual reasoning, it can be shown that the parasitic capacitance across any physical inductor will prevent an abrupt jump in the inductor voltage.

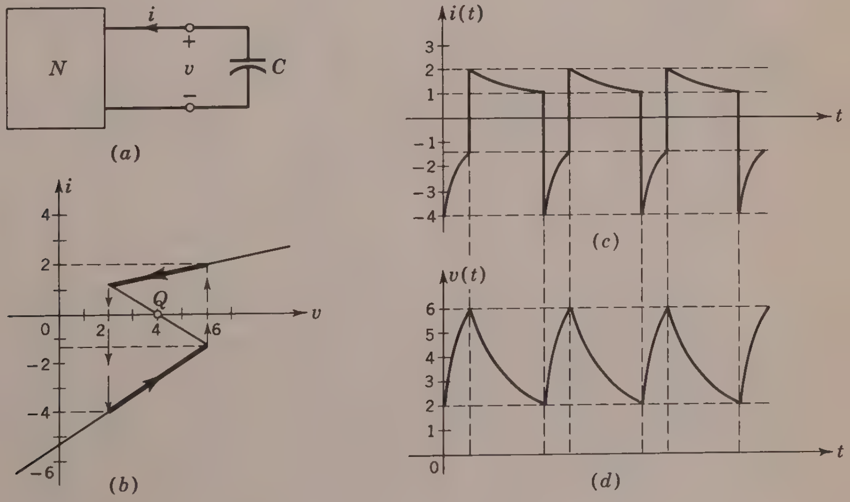
"working," the other was "relaxing." Because of this unique behavior, an astable multivibrator is sometimes called a *relaxation oscillator.*

To synthesize an astable multivibrator, the first step is to look for a resistive subnetwork which exhibits the required DP plot. Two approaches are possible: (1) One can look for commercially available nonlinear resistors whose *v-i* curves include a negative-resistance region. Some examples of these are the tunnel diode, the unijunction transistor, and the neon bulb. (2) One can apply synthesis techniques (such as those presented in Chap. 8) to realize the desired DP plot using commercially available components such as zener diodes, constant-current diodes, and transistors. Although the *v-i* curve of each of these elements does not contain a negative-resistance segment per se, a suitable interconnection of these elements can realize the desired DP plot. Since the basic principles are the same, it suffices for us to consider an illustrative example.

EXAMPLE

For our resistive subnetwork N let us choose two identical zener diodes connected back to back across a rotator as shown in Fig. 16-3*a*. The DP plot of the zener-diode combination is shown in Fig. 16-3*b*, where E_z is equal to the breakdown voltage of the zener diodes. The DP plot of the composite network N corresponding to a clockwise angle of rotation $(-90° < \theta < 0°)$ is

Fig. 16-2. The DP plot seen by the capacitor of an astable multivibrator must be current-controlled, and the negative-resistance portion of the curve must intersect the v axis.



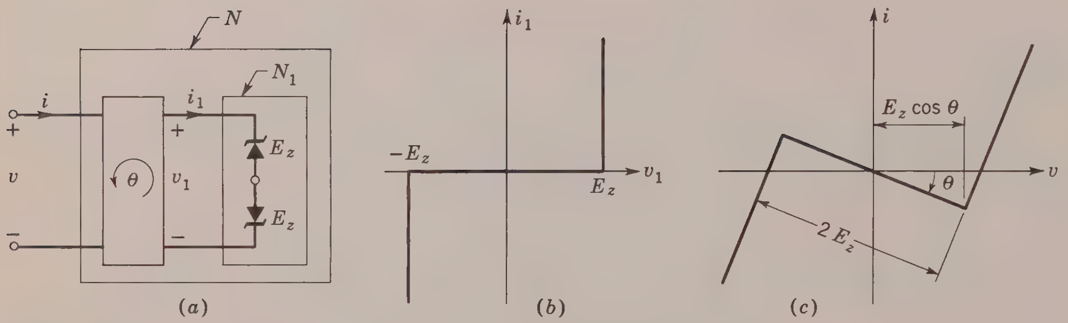


Fig. 16-3. A voltage-controlled DP plot with a negative-resistance segment intersecting the v axis can be realized by two zener diodes and a rotator.

shown in Fig. 16-3*c*. This voltage-controlled DP plot can be used to design an astable multivibrator by simply connecting an inductor L across the driving-point terminals. By a suitable choice of the value of L and the angle of rotation θ, it is possible to generate various useful periodic waveforms. Let us consider two typical cases:

Case 1: $E_z = 7.1$ volts, $L = 10$ mH, $\theta = -45°$ The network corresponding to the above choice of parameters is shown in Fig. 16-4*a*, and the DP plot seen by the inductor is shown in Fig. 16-4*b*. The resulting periodic waveform $v(t)$ is readily obtained as shown in Fig. 16-4*c*. This waveform is useful in many pulse and

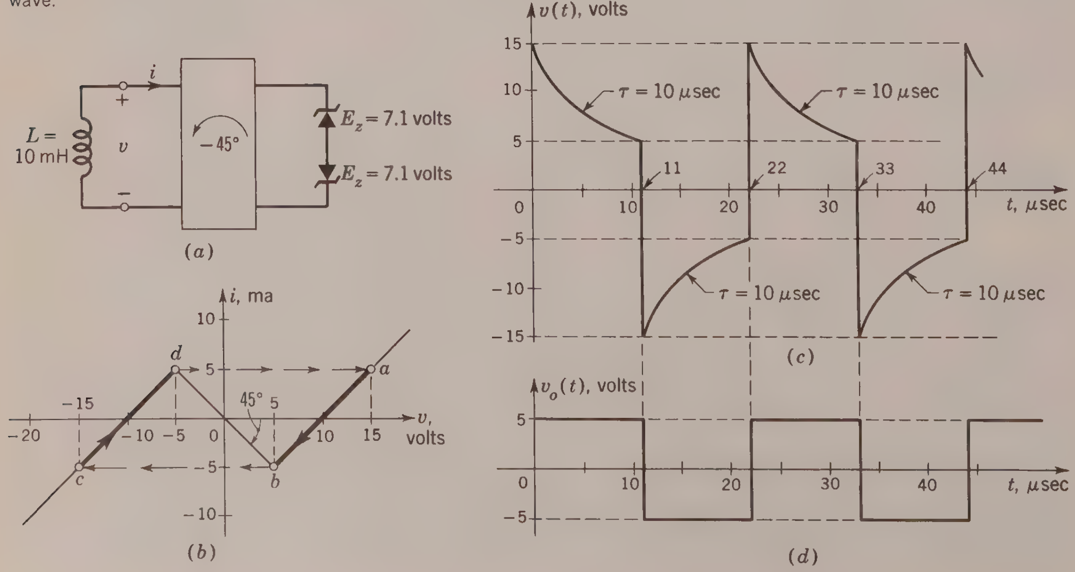


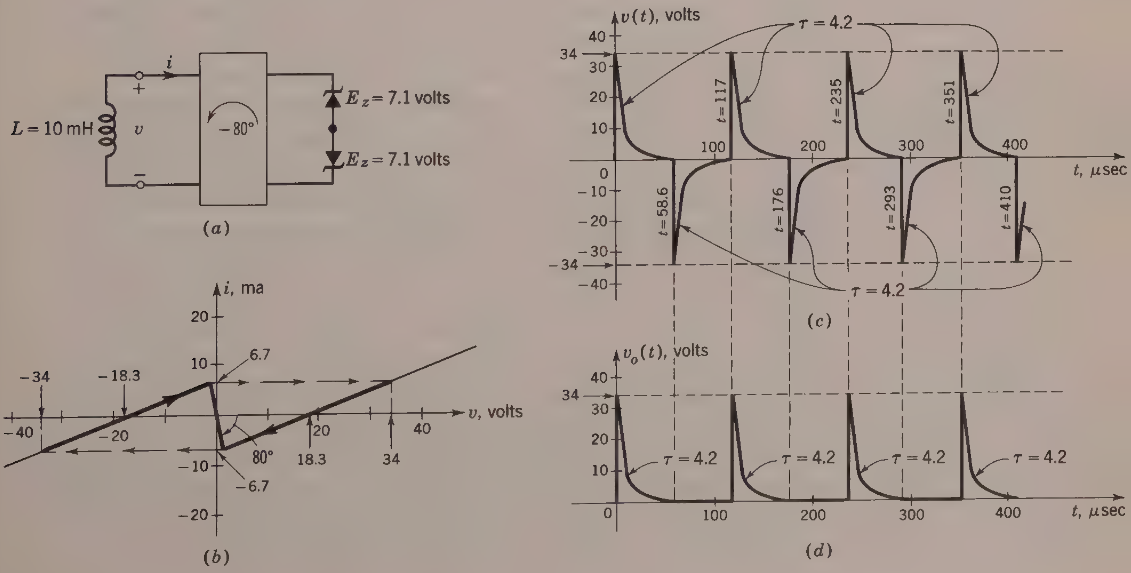
Fig. 16-4. The output waveform corresponding to case 1 consists of a periodic decaying exponential which can be clipped to obtain a square wave.

digital circuits. For example, by applying $v(t)$ across the input of a base-and-peak clipper, the square wave $v_o(t)$ shown in Fig. 16-4*d* is obtained across the output terminals of the clipper. Observe that the frequency of oscillation can be varied by changing the value of the inductance.

Case 2: $E_z = 7.1$ volts, $L = 10$ mH, $\theta = -80°$ The network corresponding to the above choice of parameters is shown in Fig. 16-5*a*, and the corresponding DP plot is shown in Fig. 16-5*b*. This time the resulting periodic waveform of $v(t)$ is as shown in Fig. 16-5*c*. Comparing this waveform with that obtained in case 1, we observe that the peak value of $v(t)$ in the present case is more than twice the peak value of $v(t)$ in case 1. Observe also that the present waveform is a lot sharper than in case 1. If we apply this waveform to the input of a base clipper, we obtain the periodic "pulse train" shown in Fig. 16-5*d*. As will be seen later, this particular waveform has widespread applications.

The above examples are chosen primarily for pedagogical reasons. They are not necessarily the best way nor the only way to generate a square wave or a pulse train under all circumstances. There are many other astable multivibrator circuits which can be used for these purposes, some of which are given in the problems.

Fig. 16-5. The output waveform corresponding to case 2 consists of a sharply decaying exponential waveform which can be clipped to obtain a periodic pulse train.



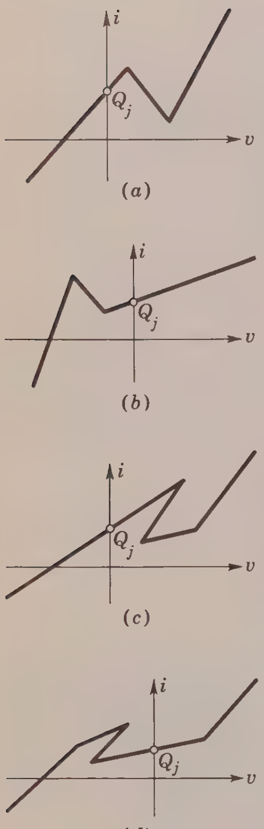
Depending on the specific applications and situations, one circuit may be preferred over another.

Exercise 1: Discuss the effect of the slope of each segment of the DP plot in Figs. 16-1*b* and 16-2*b* on the waveforms $i(t)$ and $v(t)$.

Exercise 2: Find the DP plot required to generate a 100-kHz symmetrical square wave using the following energy-storage elements: (*a*) a 100-μμF capacitor and (*b*) a 10-mH inductor.

Exercise 3: Show that an astable multivibrator can be designed by replacing the inductor and the rotator in Fig. 16-4*a* by a capacitor and a reflector, respectively. Find the range of allowable angles of reflection θ.

Exercise 4: Design a 10-kHz astable multivibrator using a type 1NXX6 tunnel diode as the nonlinear element.



(*a*)

(*b*)

(*c*)

(*d*)

Fig. 16-6. The DP plot seen by the inductor of a monostable multivibrator must be a multivalued function of current and must intersect the *i* axis at one point.

16-3 MONOSTABLE (SINGLE-SHOT) MULTIVIBRATORS

A monostable (single-shot) multivibrator is a first-order nonlinear switching network with three distinguishing properties: (1) at any time $t = t_j$ it has one, and only one, equilibrium state Q_j, (2) the equilibrium state Q_j is stable, and (3) at any time $t = t_j$ the DP plot of the resistive subnetwork is a multivalued function of current if the energy-storage element is an inductor, and a multivalued function of voltage if the energy-storage element is a capacitor.[1] For example, for an inductor the DP plots shown in Fig. 16-6*a* to *d* satisfy the above properties. Similarly, for a capacitor the DP plots shown in Fig. 16-7*a* to *d* satisfy the above properties. In either case, this multivibrator is monostable because it has *one stable* equilibrium state.

Since by definition the equilibrium state Q_j is stable, the inductor becomes a short circuit, and the capacitor becomes an open circuit at equilibrium. This is neither interesting nor useful. However, if we incorporate a triggering signal which enables us to switch instantaneously from one DP plot to another, then the resulting transient waveforms become very useful. The simplest way to incorporate a triggering signal is to connect a triggering-voltage source in series with the inductor or a triggering-current source in parallel with the capacitor as shown in Fig. 16-8*a* and *b*, respectively. Except for a horizontal or a vertical translation of the original DP plot, the form of the resulting DP plots remains unchanged. There are many other ways to connect the triggering source such that the resulting DP plot differs considerably from the original plot. However, since the basic principles are the same, we shall consider only the simpler case.

---

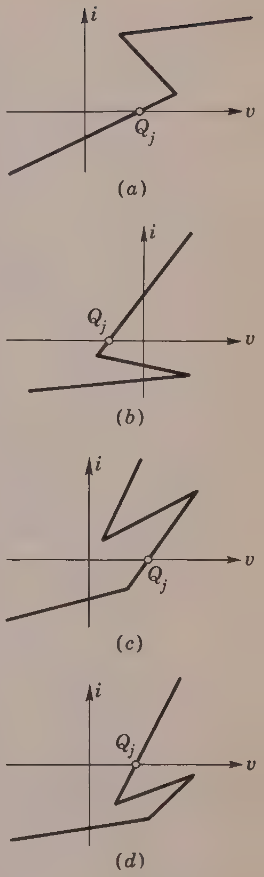
[1] Recall that a curve in the y-vs.-x plane is said to be a multivalued function of x if for some value of x there corresponds more than one value of y. It is said to be a multivalued function of y if for some value of y there corresponds more than one value of x.

EXAMPLE

The first step to synthesize a monostable multivibrator is to choose a resistive subnetwork with a suitable DP plot. An inspection of the typical DP plots shown in Fig. 16-6 might suggest a tunnel diode whose v-i curve resembles that of Fig. 16-6*a*. For simplicity, let us approximate the tunnel diode v_d-i_d curve by the piecewise-linear characteristics shown in Fig. 16-9*a*. Since this curve is voltage-controlled, an inductor must be chosen in order to satisfy the properties of a monostable multivibrator. If we connect the inductor directly across the tunnel diode, the equilibrium state will coincide with the origin. However, for reasons that will become obvious soon, it is often desirable that the equilibrium state of a monostable multivibrator be located near a breakpoint of the DP plot. We can attain this objective by connecting a resistor-battery biasing circuit in series with the inductor as shown in Fig. 16-9*b*. From the explicit equilibrium state theorem we know that the equilibrium state of this network is the same as the operating point of the resistive subnetwork obtained by short-circuiting the inductor as shown in Fig. 16-9*c*. Hence, by an appropriate choice of R and E, we can obtain a load line which intersects the v_d-i_d curve at an operating point Q near the breakpoint b, as shown in Fig. 16-9*d*. The equilibrium current is, therefore, given by $i = I_Q$.

After fixing the equilibrium state, it is convenient to obtain the composite DP plot seen by the inductor. The resulting v-i curve is simply the DP plot of the series combination of the tunnel diode, the resistor R, and the battery E. This is obtained graphically as shown in Fig. 16-9*e*. As a check, observe that at equilibrium $v = 0$, and the equilibrium current is given by $i = I_Q$. This agrees with the operating point in Fig. 16-9*d*, as it should.

The next step is to incorporate a triggering-voltage source in series with the inductor as shown in Fig. 16-10*a*. The resulting DP



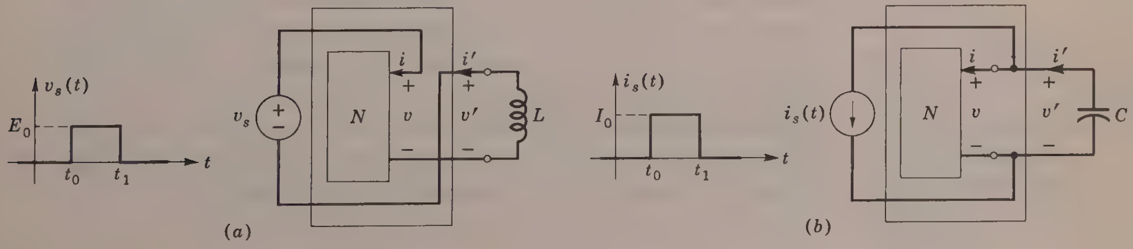
(*a*)

(*b*)

(*c*)

(*d*)

Fig. 16-7. The DP plot seen by the capacitor of a monostable multivibrator must be a multivalued function of voltage and must intersect the v axis at one point.



(*a*)

(*b*)

Fig. 16-8. Two simple methods for switching from one DP plot to another.

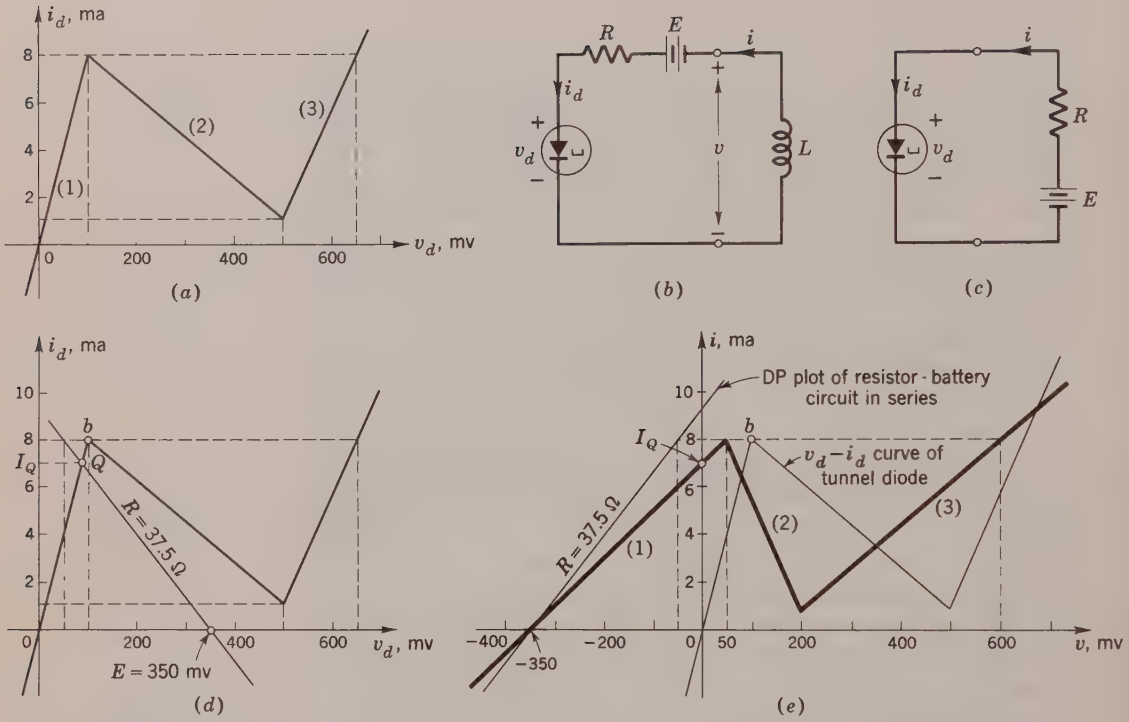


Fig. 16-9. The introduction of a resistor-battery biasing circuit in series with an inductor results in a monostable multivibrator having a prescribed equilibrium state.

plot during the triggering interval $t_0 \leq t < t_1$ is obtained by translating the original v-i curve by E_0 volts toward the left, as shown in Fig. 16-10*b*. Observe that in view of the triggering signal, the DP plot of the network switches from the DP plot I to DP plot II at $t = t_0$, and then switches back to DP plot I at $t = t_1$. The dynamic route will clearly depend on the pulse width δ of the triggering signal. A typical dynamic route suitable for a monostable multivibrator is shown in Fig. 16-10*b*. Here we assume that the dynamic route reaches point *d* at $t = t_1^-$. At $t = t_1^+$ the DP plot switches instantaneously back to DP plot I. Since the inductor current cannot change instantly, the dynamic route must be along the horizontal line $i = I_d$. Since there are three possible candidates (e, e', e''), we must invoke the inertia postulate to conclude that the dynamic route must go to point *e* at $t = t_1^+$.†

The remaining portions of the dynamic route are easily obtained as shown in Fig. 16-10*b*. The corresponding output waveforms of $i(t)$ and $v_L(t)$ are shown in Fig. 16-11*a* and *b*.

---

† Observe that at $t = t_1^-$ the network is operating on branch 3 of the tunnel diode. Therefore, at $t = t_1^+$ it must remain on branch 3, if at all possible. This is the essence of the inertia postulate.

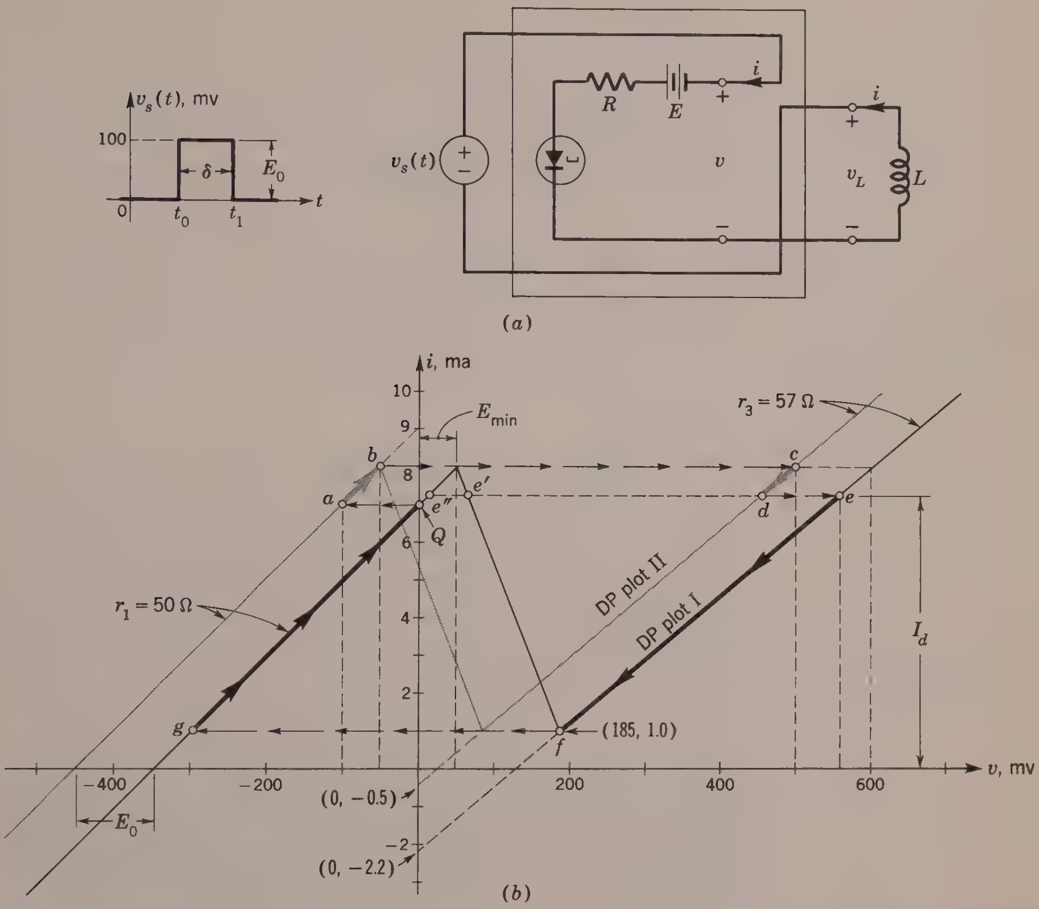


Fig. 16-10. The dynamic route of a monostable multivibrator during the triggering operation.

16-3-1 SOME APPLICATIONS OF MONOSTABLE MULTIVIBRATORS

Most applications of monostable multivibrators are based on two significant properties of the output voltage waveform $v(t)$.† The first property concerns the observation that the peak value of the output voltage ($v_L = 560$ mv) is many times larger than the triggering amplitude ($E_0 = 100$ mv). Moreover, since this peak output voltage is independent of the triggering amplitude, a monostable multivibrator can be used to produce a large output voltage signal whenever a relatively small input pulse is present. In this sense, the monostable multivibrator is used as a pulse amplifier. An example of this application is the detection of nuclear radiation by a Geiger counter, which operates on the principle that a weak

---

† For a capacitor it is the output current waveform that is useful.

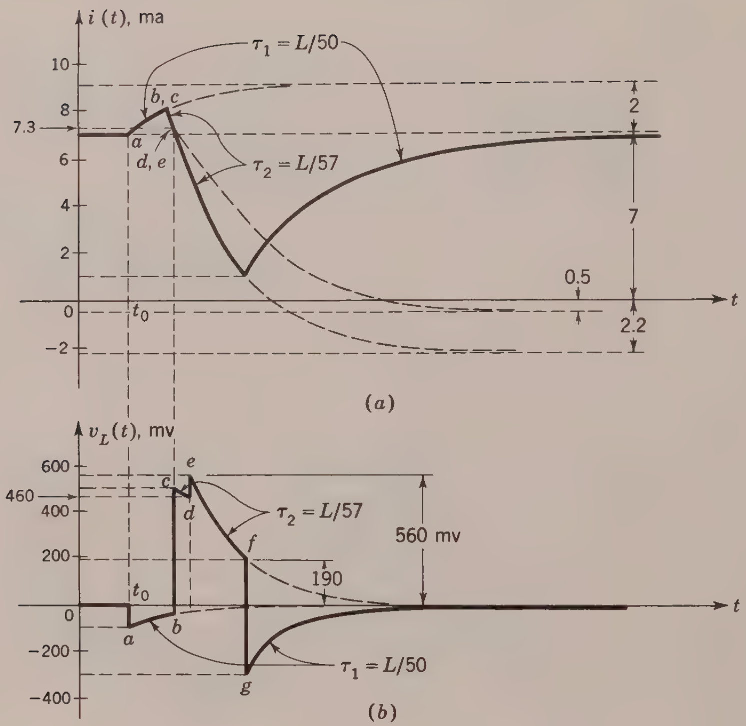


Fig. 16-11. The output waveforms of a monostable multivibrator during triggering operation.

current pulse is generated whenever a radiation particle passes through the counter tube. This weak current pulse can be used to trigger a monostable multivibrator, thus generating an output pulse large enough to deflect an indicating instrument such as a galvanometer.

The second property concerns the abrupt-jump phenomenon of the output waveform $v(t)$. If this signal is applied to a base-and-peak clipper, the square output pulse $v_o(t)$ shown in Fig. 6-12*a* is obtained. Observe that this pulse starts at $t = t_b$ where $t_b > t_0$. If we draw the triggering pulse $v_s(t)$ directly below $v(t)$ as shown in Fig. 16-12*b*, we find that the output pulse is delayed by $\Delta = t_b - t_0$ sec. Therefore, a monostable multivibrator can be designed to function as a "delay line," an indispensable building block in many electronic systems. By an appropriate choice of the time constant $\tau_1 = L/r_1$, any desired delay Δ can be obtained.

So far our discussions are based on the dynamic route shown in Fig. 16-10*b*. Let us now investigate the effects of the amplitude E_0 and the pulse width δ of the triggering signal on the dynamic route and the associated output voltage $v(t)$.

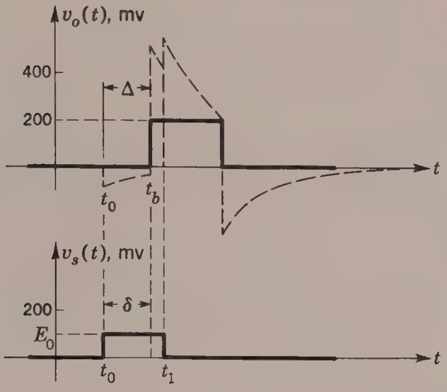
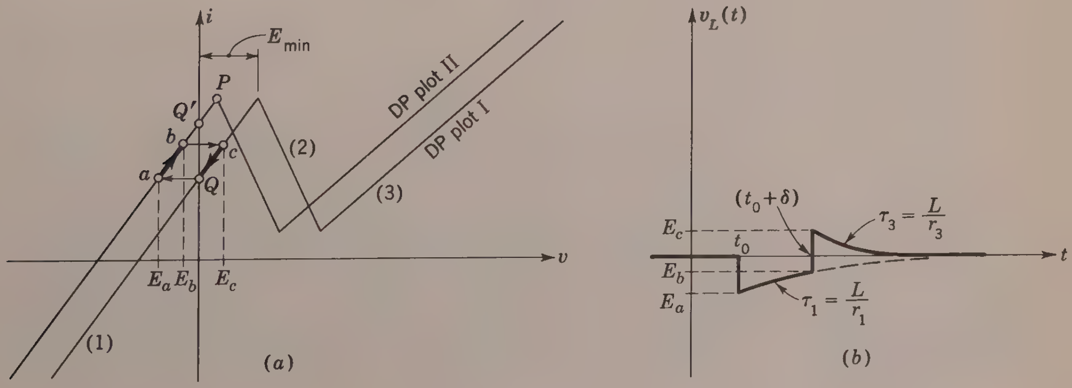


Fig. 16-12. An illustration showing how the output voltage of a monostable multivibrator can be clipped to obtain a delayed square pulse.

16-3-2 MINIMUM TRIGGERING-PULSE-AMPLITUDE CONDITION

Consider first the case where $E_0 < E_{\min}$, where $E_{\min}$ is the voltage at the upper breakpoint of the DP plot shown in Fig. 16-10*b*. The resulting DP plot II during $t_0 \leq t < t_1$ is shown in Fig. 16-13*a*. Observe that since the breakpoint P of DP plot II is in the right half-plane, the dynamic route will be as shown in Fig. 6-13*a*. Clearly, this dynamic route will never reach point P because it will take an infinite time interval to arrive even at the new equilibrium state Q'. Therefore, the dynamic route will arrive at some point b at $t = t_0 + \delta$ and then jump back to DP plot I. The corresponding waveform for $v_L(t)$ is shown in Fig. 16-13*b*. Observe that the desirable property of a large abrupt jump is no longer present. In fact, the magnitude of the jump in $v_L(t)$ is equal to E_0, the triggering amplitude! Since nothing is gained here, this situation is undesirable, and the multivibrator is said to have "failed to trigger."

Fig. 16-13. The monostable multivibrator fails to trigger whenever $E_0 < E_{\min}$.



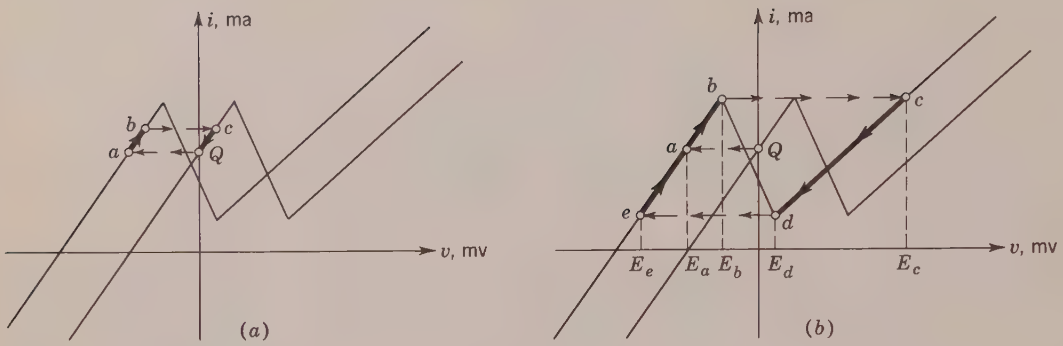


Fig. 16-14. The dynamic route on the left occurs whenever the trigger-pulse width is too small. The one on the right occurs whenever the trigger-pulse width is too large.

In order to avoid this situation, an inspection of Fig. 16-13*a* shows that the triggering amplitude should be chosen large enough so that the upper breakpoint *P* of DP plot II will clear the *i* axis and move into the left half-plane. The condition to satisfy this requirement is clearly given by

$$E_0 > E_{\min} \tag{16-1}$$

16-3-3 MINIMUM TRIGGERING-PULSE-WIDTH CONDITION

Consider next the effect of the pulse width δ on $v_L(t)$, assuming that Eq. (16-1) is satisfied. An inspection of the dynamic route of Fig. 16-10*b* shows that if δ is smaller than the time it takes the route to arrive at the peak point of DP plot II ($\delta < t_b - t_0$), then the dynamic route will be as shown in Fig. 16-14*a*. Since this is similar to the route of Fig. 16-13*a*, the multivibrator will fail to trigger. Therefore, for satisfactory triggering to occur, the pulse width δ must be larger than the time $\delta_{\min}$ it takes the dynamic route to clear point *P* of DP plot II, or

$$\delta > \delta_{\min} \tag{16-2}$$

What happens if δ is too large? The corresponding dynamic route shown in Fig. 16-14*b* will repeat itself for as long as the trigger pulse is on. Since the circuit becomes an astable multivibrator for all times $t_0 \leq t < t_1$, the output voltage $v_L(t)$ will consist of several cycles of oscillations as shown in Fig. 16-15.[1] This waveform is undesirable in the application described above. However, it is useful in other applications, such as in the design of a "burst generator," where a periodic "burst" of a few cycles of the output signal is desired. Another application where a gated oscillator may be useful is in the design of a threshold device for generating a persistent warning signal whenever the input signal reaches a predetermined "dangerous" level.

---

[1] A circuit with this property is usually called a *gated oscillator.*

Exercise 1: Verify that the multivalued DP plot shown in Fig. 16-6*c* and *d* can be synthesized by connecting a nonlinear resistor with an appropriate voltage-controlled DP plot in series with a positive linear resistor.

Exercise 2: Verify that the multivalued DP plot shown in Fig. 16-7*c* and *d* can be synthesized by connecting a nonlinear resistor with an appropriate current-controlled DP plot in parallel with a positive linear resistor.

Exercise 3: Derive a systematic procedure for determining the minimum triggering-pulse amplitude and the minimum triggering-pulse width directly from the load-line construction shown in Fig. 16-9*d*.

Exercise 4: Discuss the triggering mechanisms for a first-order monostable multivibrator with a linear capacitor as the energy-storage element. Specify the conditions on the triggering-pulse amplitude and triggering-pulse width for successful operation.

Exericse 5: Enumerate some practical applications of a delay line and a gated oscillator.

16-4 BISTABLE (DOUBLE-SHOT, FLIP-FLOP) MULTIVIBRATORS

A bistable (double-shot, flip-flop) multivibrator is a first-order nonlinear switching network with three distinguishing properties: (1) it has three equilibrium states Q_1, Q_2, and Q_3, (2) at least two equilibrium states are stable, and (3) the DP plot of the resistive subnetwork is a multivalued function of voltage if the energy-

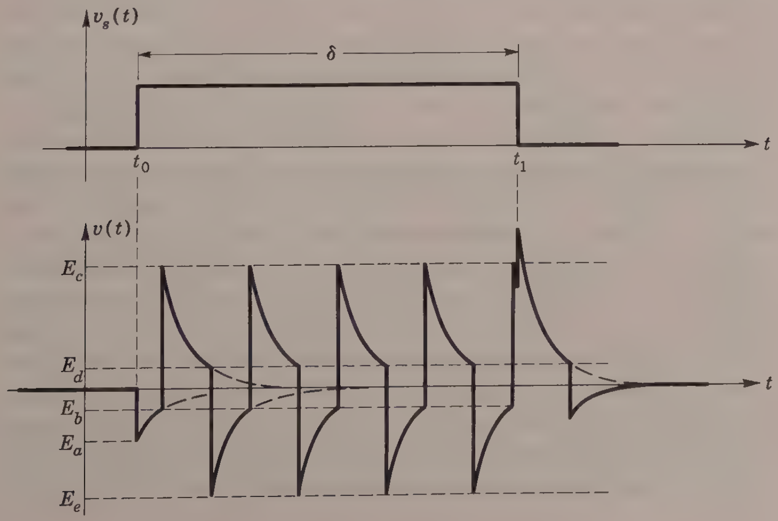


Fig. 16-15. The monostable multivibrator becomes a gated oscillator when the trigger pulse width is too large.

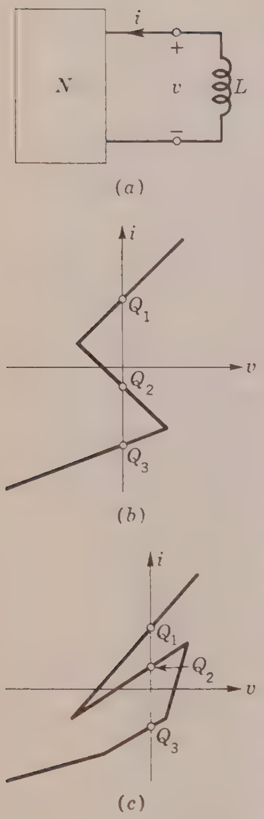


Fig. 16-16. The DP plot seen by the inductor of a bistable multivibrator must be a multivalued function of voltage and must intersect the i axis at three points.

storage element is an inductor, and a multivalued function of current if the energy-storage element is a capacitor. For example, for an inductor (Fig. 16-16*a*) both the current-controlled DP plot of Fig. 16-16*b* and the multivalued DP plot of Fig. 16-16*c* are multivalued functions of voltage and are, therefore, possible candidates for a bistable multivibrator. For a capacitor (Fig. 16-17*a*) both the voltage-controlled DP plot of Fig. 16-17*b* and the multivalued DP plot of Fig. 16-17*c* are multivalued functions of current and are, therefore, possible candidates. Observe that two of the three equilibrium states Q_1 and Q_3 in Fig. 16-16*b* and 16-17*b* are stable, whereas the third equilibrium state Q_2 is unstable. Observe also that all three equilibrium states in Figs. 16-16*c* and 16-17*c* are stable. However, the equilibrium state Q_2 in this case is not "accessible" in the sense that unless the initial condition was originally located in that branch of the DP plot containing this equilibrium state, it is not possible to reach Q_2.† From experience we know that the initial state is never located in this branch, and hence even though Q_2 is stable, it is never observed. We can, therefore, conclude that the DP plots shown in Figs. 16-16 and 16-17 have two accessible stable equilibrium states, hence the name bistable.

The two possible equilibrium states of a bistable multivibrator can be used to represent the two distinct states, usually labeled "state 0" and "state 1," in binary logic. For example, whenever the multivibrator is operating at Q_1, we can agree to call it state 0 and whenever the multivibrator is operating at Q_3, we can agree to call it state 1. This provides a simple form of "memory." By a systematic arrangement of many bistable multivibrators, it is possible to store information in digital form. It is, therefore, not surprising that bistable multivibrators are the building blocks of many digital computers. In order to be able to store the appropriate information, a means must be provided so that one can force a bistable multivibrator to operate at either of the two equilibrium states. This can be achieved by a triggering source similar to that used in triggering a monostable multivibrator. Before we present a detailed example, it is instructive to consider a mechanical analogy consisting of a glass tube (closed on both ends) with a metallic ball inside it as shown in Fig. 16-18*a*. The tube is free to turn about the pivot on top of the wedge, and the ball will move along the tube in accordance with the laws of mechanics. Observe that of the three equilibrium states Q_1, Q_2, and Q_3 as shown in Fig. 16-18*b* to *d*, respectively, only Q_1 and Q_3 are stable. If we denote the position Q_1 by 0 and the position Q_3 by 1, we obtain a

---

† In view of the inertia postulate the dynamic route can jump only after reaching a breakpoint. Hence, if the initial state is along the segment containing either Q_1 or Q_3, the dynamic route will never reach Q_2.

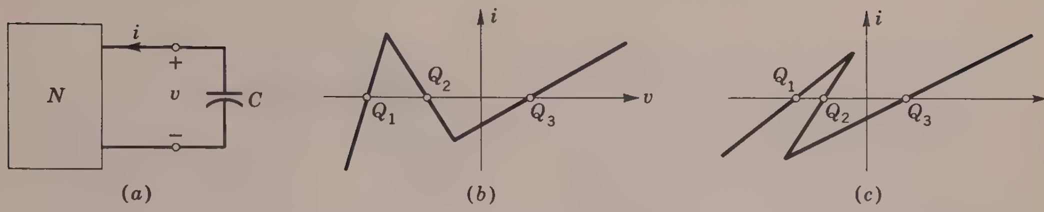
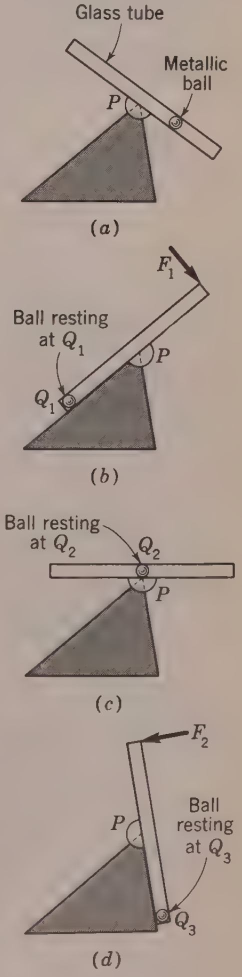


Fig. 16-17. The DP plot seen by the capacitor of a bistable multivibrator must be a multivalued function of current and must intersect the v axis at three points.

Fig. 16-18. The mechanical analogy of a flip-flop.



two-state device similar to a bistable multivibrator. To carry the analogy further, observe that we can trigger from state 0 to state 1 by applying a force F_1 to "flip" the glass tube (Fig. 16-18*b*) until it "flops" over to the other side of the wedge (Fig. 16-18*d*). In view of this switching mechanism, a bistable multivibrator is commonly called a flip-flop. Observe that in order to trigger successfully from state 0 to state 1, a minimum force F_1 is required to overcome the weight of the ball. Moreover, this force must be applied for a minimum period equal to the time it takes the ball to move past the pivoting point P as shown in Fig. 16-18*b*. Similarly, a minimum triggering force F_2 must be applied (Fig. 16-17*d*) over a minimum period in order to trigger from state 1 back to state 0. Let us now consider the design of an electronic flip-flop.

EXAMPLE

For convenience, let us choose a tunnel diode again, but this time we shall use it to synthesize a bistable multivibrator. In view of the explicit equilibrium state theorem, it is clear that in order to obtain three equilibrium states, the resistive network obtained by short-circuiting the inductor or open-circuiting the capacitor must have three operating points. Hence the first step is to synthesize a biasing network so that its load line intersects the v_d-i_d curve of the tunnel diode at three points. For reasons that will be obvious soon, it is desirable that these points be located near the breakpoints of the v_d-i_d curve. One simple biasing network for accomplishing this is shown in Fig. 16-19*a*. The corresponding operating points are shown in Fig. 16-19*b*. The next step is to select the energy-storage element and to connect it in such a way that the corresponding resistive network in equilibrium reduces to that shown in Fig. 16-19*a*. Since an inductor becomes a short circuit and a capacitor becomes an open circuit when the network is in equilibrium, it is clear that if an inductor is chosen, it must be connected in series with the biasing network as shown in Fig.

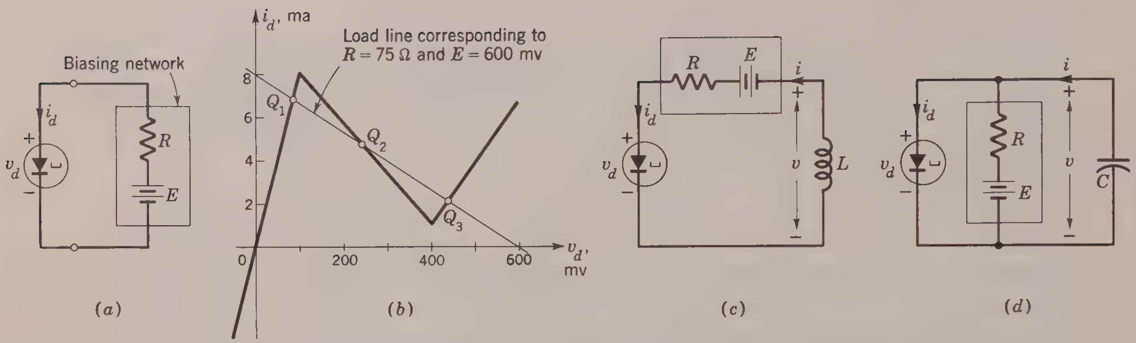
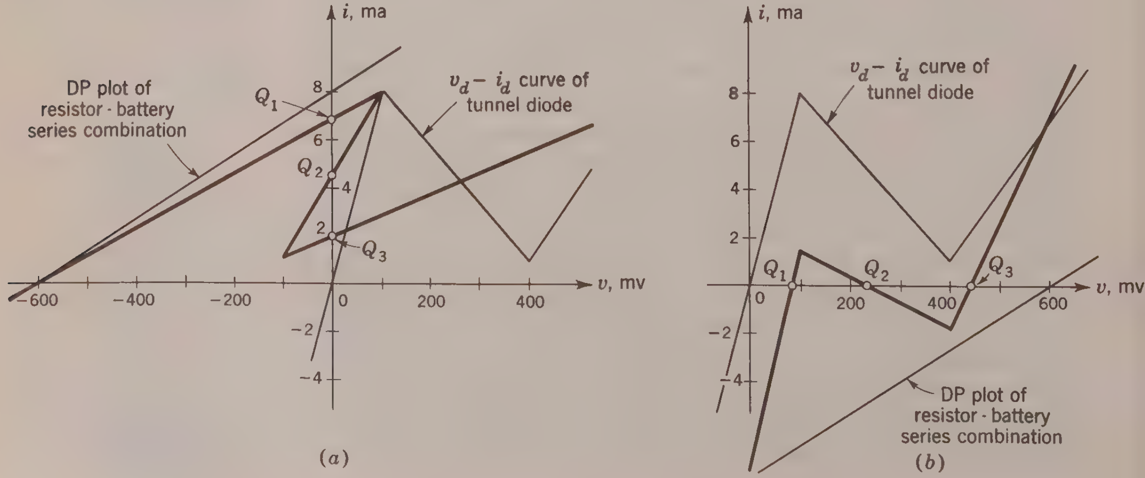


Fig. 16-19. The first step to synthesize a flip-flop is to design a biasing network which yields three operating points. The next step is to choose the energy-storage element and the appropriate connections.

16-19*c*. Similarly, if a capacitor is chosen, it must be connected in parallel with the biasing network as shown in Fig. 16-19*d*. To verify that these two basic network configurations indeed satisfy the properties defining a bistable multivibrator, the DP plot as seen by the inductor is shown in Fig. 16-20*a*, and the DP plot as seen by the capacitor is shown in Fig. 16-20*b*. In the first case the equilibrium states Q_1, Q_2, and Q_3 are found at the intersections of the multivalued v-i curve with the i axis. In the second case the equilibrium states Q_1, Q_2, and Q_3 are found at the intersections of the voltage-controlled v-i curve with the v axis. In this case observe that Q_1 and Q_3 are stable but Q_2 is unstable. It remains for us to design an appropriate triggering circuit so that we can switch from one state to another.

Case 1: Triggering the inductor flip-flop The simplest way to trigger the inductor flip-flop circuit of Fig. 16-19*c* is to connect a voltage

Fig. 16-20. The DP plot seen by the inductor is a multivalued function of voltage, and the DP plot seen by the capacitor is a multivalued function of current.



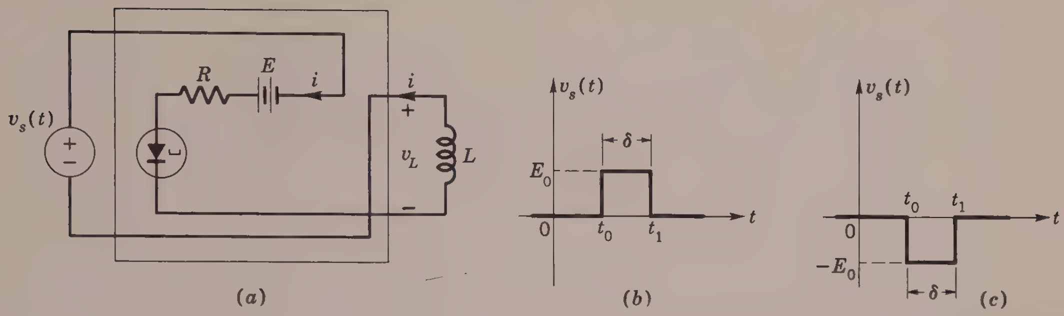


Fig. 16-21. The simplest triggering circuit for an inductor flip-flop consists of a positive or a negative square-voltage-pulse triggering source in series with the inductor.

source in series with the inductor as shown in Fig. 16-21*a*. Depending on the initial equilibrium state, the triggering voltage may be either a positive or a negative square pulse as shown in Fig. 16-21*b* and *c*, respectively. If the initial state is at Q_1, it is necessary to apply a positive pulse to switch from Q_1 to Q_3. If the initial state is at Q_3, it is necessary to apply a negative pulse to switch from Q_3 to Q_1. The pertinent DP plots before and after the triggering pulse is applied are shown in Fig. 16-22*a* corresponding to a positive pulse and in Fig. 16-22*b* corresponding to a negative pulse. Also shown in these figures are the dynamic routes corresponding to a typical pulse height E_0 and pulse width δ. For illustrative purposes, the output waveforms corresponding to the triggering signal shown in Fig. 16-23*a* are shown in Fig. 16-23*b* and *c*.

Let us now investigate the effects of the triggering-pulse height E_0 and pulse width δ on the triggering operation. It suffices to consider the case where we want to switch from Q_1 to Q_3. If the pulse height E_0 is not chosen large enough for the upper breakpoint of the shifted DP plot to clear the *i* axis, then the dynamic route will be as shown in Fig. 16-24*a*. Observe that this route

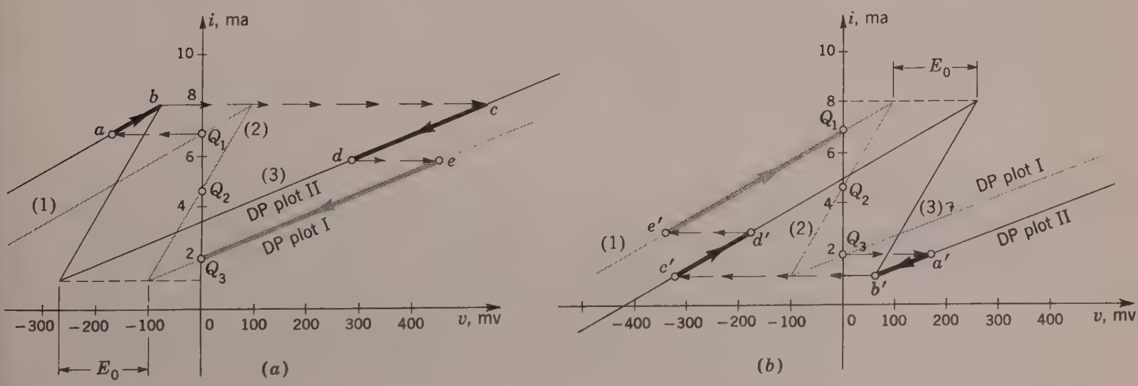


Fig. 16-22. The dynamic routes corresponding to a typical positive triggering-voltage pulse and a negative triggering-voltage pulse.

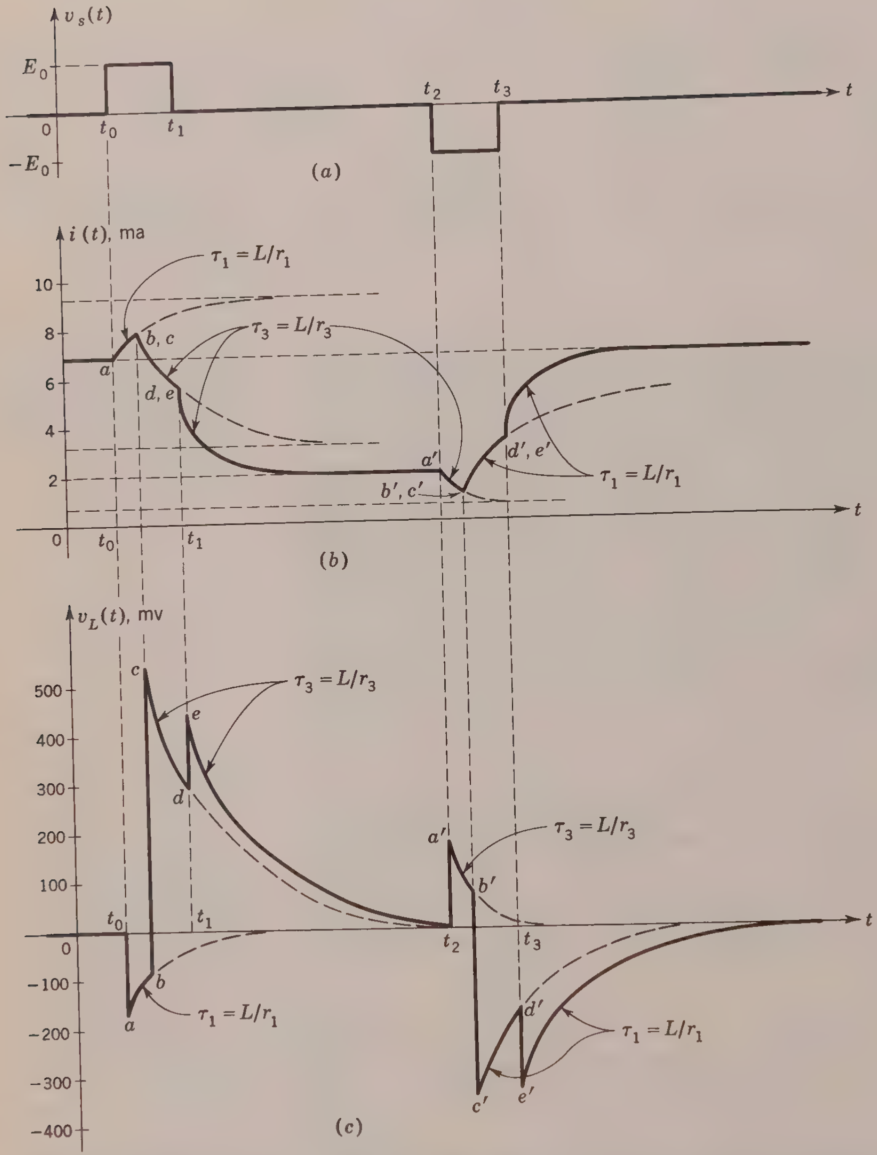


Fig. 16-23. The output waveforms corresponding to a positive and a negative triggering-voltage pulse applied in succession.

eventually returns to Q_1, and the flip-flop fails to trigger. Clearly, in order to trigger successfully, the value of E_0 must exceed the value of $E_{\min}$ shown in Fig. 16-24*a*. Consider next what happens if the pulse width is too short. The resulting dynamic route in Fig. 16-24*b* shows that once again the operation returns to Q_1, and the flip-flop fails to trigger. Clearly, the minimum value

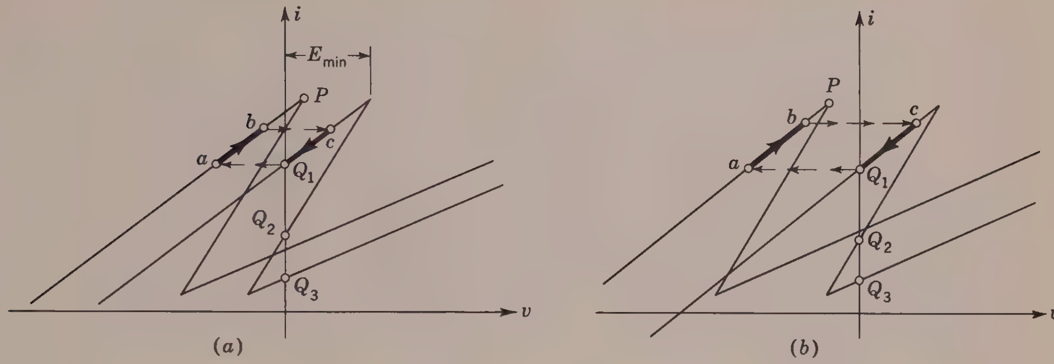
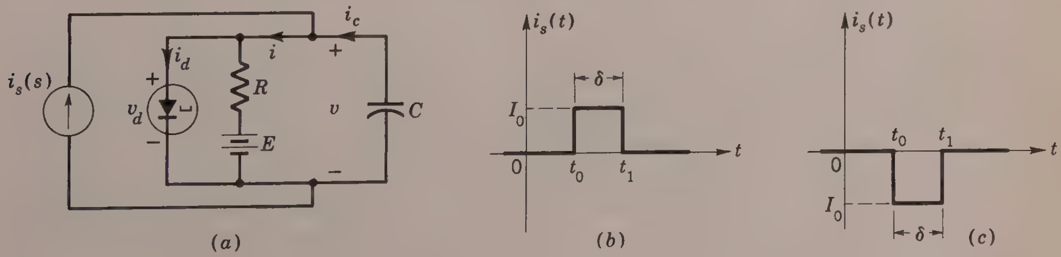


Fig. 16-24. The flip-flop will fail to trigger if the pulse height E_0 or the pulse width δ is too small.

$\delta_{\min}$ of the pulse width must be equal to the time it takes the route to arrive at the upper breakpoint P shown in Fig. 16-24*b*.

Case 2: Triggering the capacitor flip-flop The simplest way to trigger the capacitor flip-flop of Fig. 16-19*d* is to connect a current source in parallel with the capacitor as shown in Fig. 16-25*a*. The positive-current pulse shown in Fig. 16-25*b* is required to switch from Q_1 to Q_3. Conversely, the negative-current pulse shown in Fig. 16-25*c* is required to switch from Q_3 to Q_1. The pertinent DP plots before and after the triggering pulse is applied are shown in Fig. 16-26*a* corresponding to a positive pulse and in Fig. 16-26*b* corresponding to a negative pulse. Also shown in these figures are the dynamic routes corresponding to a typical pulse height I_0 and pulse width δ. For illustrative purposes, the output waveforms corresponding to the triggering signal shown in Fig. 16-27*a* are shown in Fig. 16-27*b* and *c*. It is easy to see that if I_0 is too small, the dynamic route will be as shown in Fig. 16-28*a*, and the flip-flop fails to trigger. Similarly, if δ is too small, the dynamic route will be as shown in Fig. 16-28*b*, and the flip-flop also fails to trigger. To trigger successfully, it is necessary that $I_0 > I_{\min}$ and $\delta > \delta_{\min}$, where $I_{\min}$ is the peak current shown in Fig. 16-26*a* and $\delta_{\min}$ is the

Fig. 16-25. The simplest triggering circuit for a capacitor flip-flop consists of a positive or a negative square-current-pulse triggering source in parallel with the capacitor.



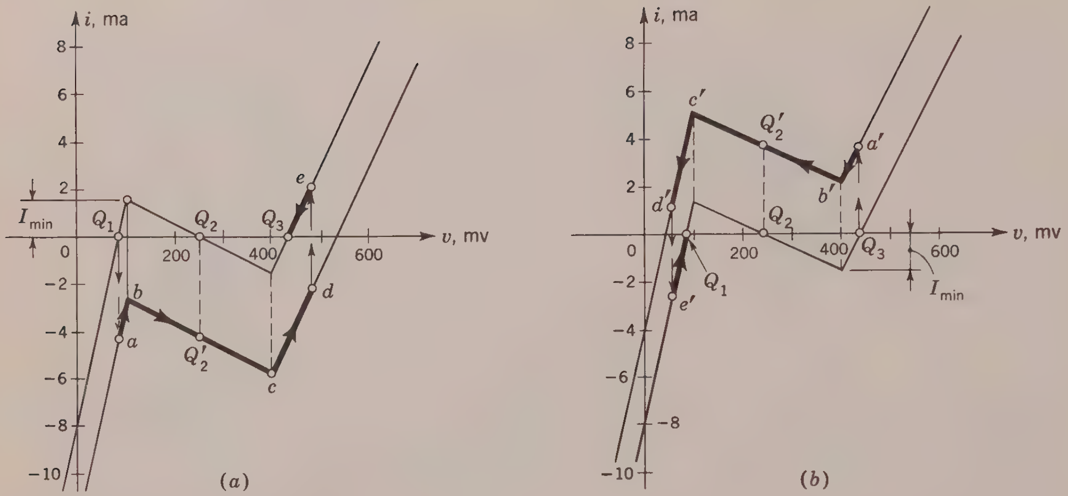


Fig. 16-26. The dynamic routes corresponding to a typical positive triggering-current pulse and a negative triggering-current pulse.

time it takes the dynamic route to reach the point Q_2' shown in Fig. 16-28*b*.

From the output waveforms shown in Figs. 16-23 and 16-27, we observe that for the inductor flip-flop the output current $i(t)$ is the useful quantity because it exhibits two distinct values and can, therefore, be used to represent the two states of the binary logic. On the other hand, for the capacitor flip-flop the output voltage $v(t)$ is the useful quantity. Observe that in both cases an additional period of time after the triggering signal has dropped to zero is needed for the respective output variable to reach the new equilibrium value. Since the speed of a digital computer is a function of how fast a flip-flop can be triggered from one state to another, it is highly desirable that this transient time be as short as possible. Since this depends on the time constant L/r_j or r_jC, the transient time can be minimized by choosing as small a value for L or C as possible. It is not possible to reduce L or C to zero because some parasitic inductance and capacitance are invariably present in any physical circuit. On the other hand, there is no reason for adding an external inductance to the inductor flip-flop or an external capacitance to the capacitor flip-flop because this will only increase the time constant of the resulting network. In other words, we need not use any physical inductance or capacitance to build a practical flip-flop because the parasitic inductance or the parasitic capacitance will assume the role of the desired energy-storage element. This is why most commercial flip-flop circuits do not show any energy-storage element explicitly.

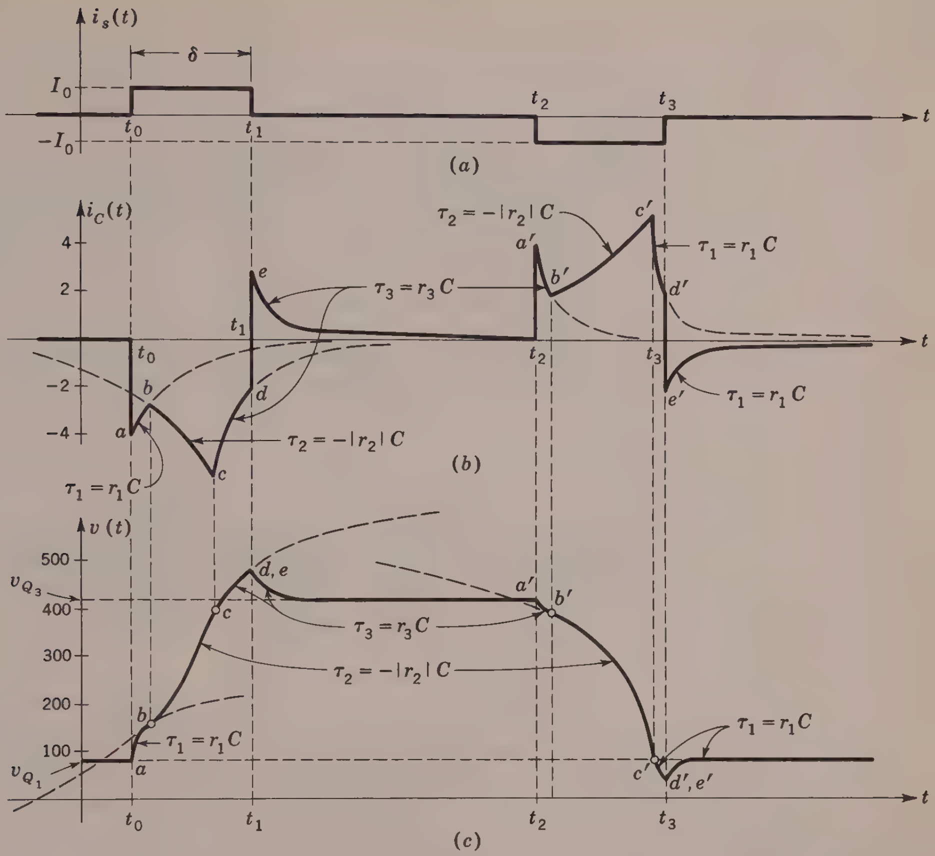


Fig. 16-27. The output waveforms corresponding to a positive and a negative triggering-current pulse applied in succession.

The above discussion is based on the assumption that only one parasitic element plays a major role. Strictly speaking, however, there is more than one parasitic element, and hence a flip-flop is actually an *n*th-order network. Fortunately, the error resulting from assuming only one of these parasitic elements is usually quite small for most practical purposes.

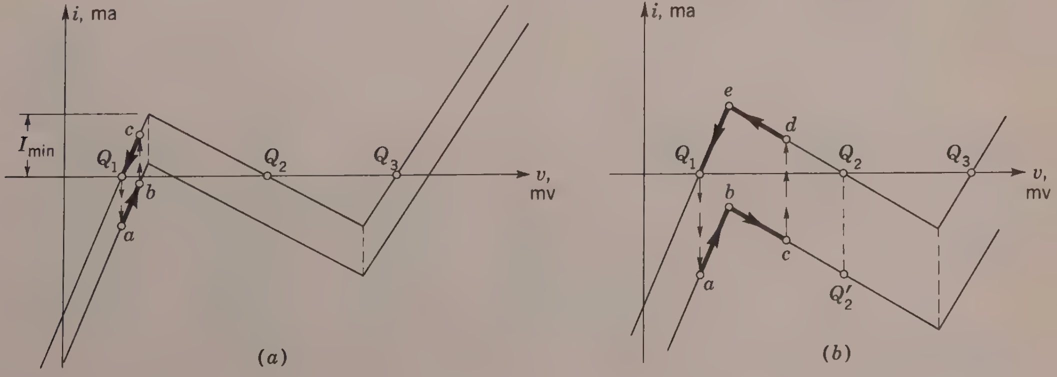


Fig. 16-28. The flip-flop will fail to trigger if the current pulse height I_0 or pulse width δ is too small.

Exercise 1: Compare the advantages and disadvantages of the two bistable multivibrator circuits shown in Figs. 16-21 and 16-25.

Exercise 2: Design a bistable multivibrator using a type 1NXX8 glow lamp and a linear capacitor as the energy-storage element. Specify the conditions on the triggering-pulse amplitude and the triggering-pulse width for successful operation.

Exercise 3: Repeat Exercise 1 with a linear inductor as the energy-storage element.

Exercise 4: Connect a small parasitic capacitor across terminals *a-b* of the SCR in the latching switch shown in Fig. 10-6*f* and explain the switching mechanisms. Show that the presence of this parasitic capacitor prevents an instantaneous switching. Sketch the transient waveforms.

16-5 SUMMARY

First-order multivibrator A first-order dynamic network characterized by the property that either the voltage or the current waveform exhibits an "abrupt jump" either periodically or whenever a relatively small triggering signal is applied.

Three types of first-order multivibrators

1. Astable (Free-running) Multivibrator
2. Monostable (Single-shot) Multivibrator
3. Bistable (Double-shot, Flip-flop) Multivibrator

TABLE 16-1 *Properties of first-order multivibrators.*

	Astable		Monostable		Bistable	
Defining properties	Inductor-loaded	Capacitor-loaded	Inductor-loaded	Capacitor-loaded	Inductor-loaded	Capacitor-loaded
Number of equilibrium states	1	1	1	1	3	3
Type of equilibrium state	Unstable	Unstable	Stable	Stable	The two outermost equilibrium states must be stable[1]	The two outermost equilibrium states must be stable[1]
Type of DP plot of the resistive subnetwork	Voltage-controlled	Current-controlled	Voltage-controlled or multivalued function of current	Current-controlled or multivalued function of voltage	Current-controlled or multivalued function of voltage	Voltage-controlled or multivalued function of current

[1] The center equilibrium state may be either stable or unstable.

PROBLEMS

16-1 The DP plots of the resistive portion of an adjustable waveform generator, used in an automatic milling-machine controller, are shown in Fig. P16-1*a* to *c*. The controller sequence requires that a capacitor or an inductor be connected across the terminals of the resistive black box as shown in Fig. P16-1*d* and *e*.

(a) Plot the waveforms of $v_C(t)$ and $i(t)$ for the circuit of Fig. P16-1*d* and the DP plot of Fig. P16-1*a* when $v_C(0) = -10$ volts.

(b) Plot the waveforms of $v_C(t)$ and $i(t)$ for the circuit of Fig. P16-1*d* and the DP plot of Fig. P16-1*b* when $v_C(0) = -10$ volts.

(c) Plot the waveforms of $v_L(t)$ and $i(t)$ for the circuit of Fig. P16-1*e* and the DP plot of Fig. P16-1*b* when $i_L(0) = -3$ ma.

(d) Plot the waveforms of $v_C(t)$ and $i(t)$ for the circuit of Fig. P16-1*d* and the DP plot of Fig. P16-1*c* when $v_C(0) = -10$ volts.

(e) Classify the four circuit combinations above according to whether the operation is astable, monostable, or bistable.

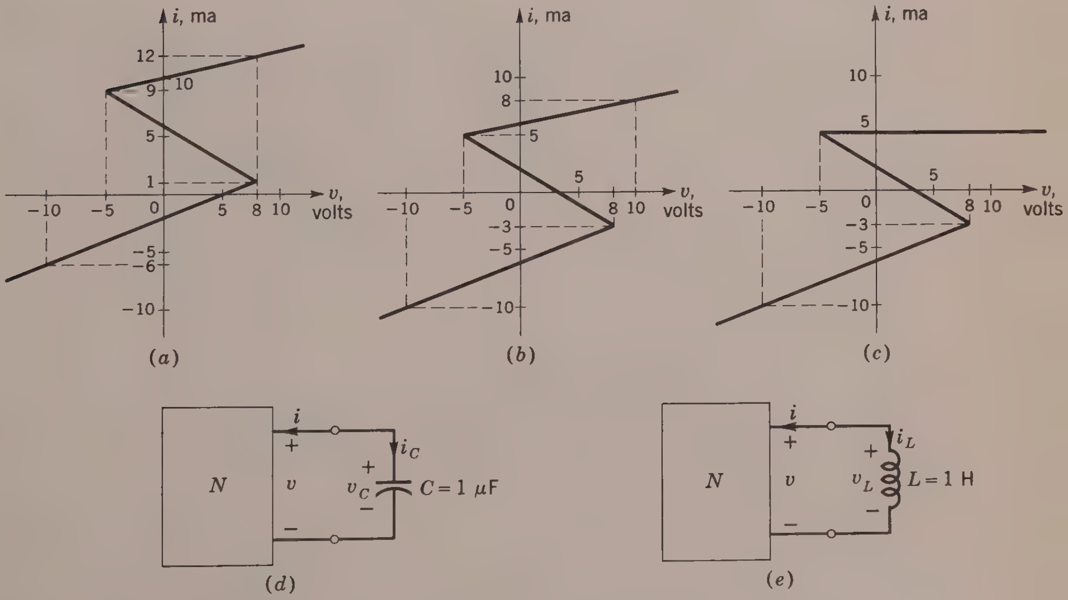


Fig. P16-1.

16-2 Consider the tunnel-diode circuit shown in Fig. P16-2*a* and the v_d-i_d curve shown in Fig. P16-2*b*.

(a) If $R = 100\ \Omega$, find the range of E for monostable operation and the range of E for bistable operation.

(b) If $R = 100\ \Omega$ and $E = 450$ mv, find the minimum amplitude of a rectangular voltage pulse required to trigger the circuit. Show how the pulse is to be applied.

(c) If $R = 100\ \Omega$, $E = 450$ mv, and $L = 15$ mH, find the minimum pulse width of a 200-mv triggering pulse for successful operation.

(d) If $R = 100\ \Omega$, $E = 450$ mv, and $L = 15$ mH, find $v_d(t)$ and $i_d(t)$ corresponding to a 200-mv triggering pulse with a pulse width of 250 μsec.

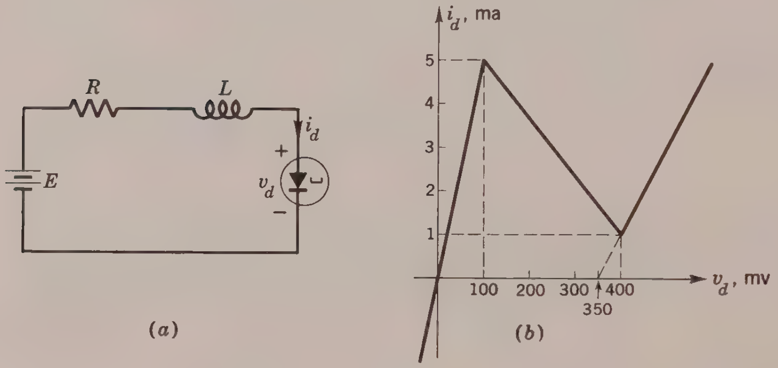


Fig. P16-2.

16-3 Consider the flip-flop circuit shown in Fig. P16-3*a* where the DP plot of the resistive black box *N* is given by Fig. P16-3*b*.

(a) If the triggering signal $v_s(t)$ is as shown in Fig. P16-3*c* with $E_0 = 2$ volts, find the minimum pulse width δ required to trigger the flip-flop from its upper state to its lower state.

(b) If $\delta = 0.5$ msec, find the minimum pulse amplitude E_0 required to trigger the flip-flop from its upper state to its lower state.

(c) If $E_0 = 2$ volts and $\delta = 0.5$ msec, find $v_L(t)$ and $i(t)$.

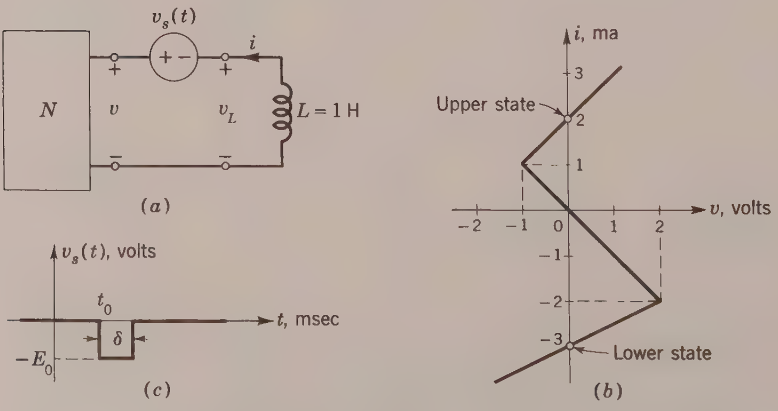
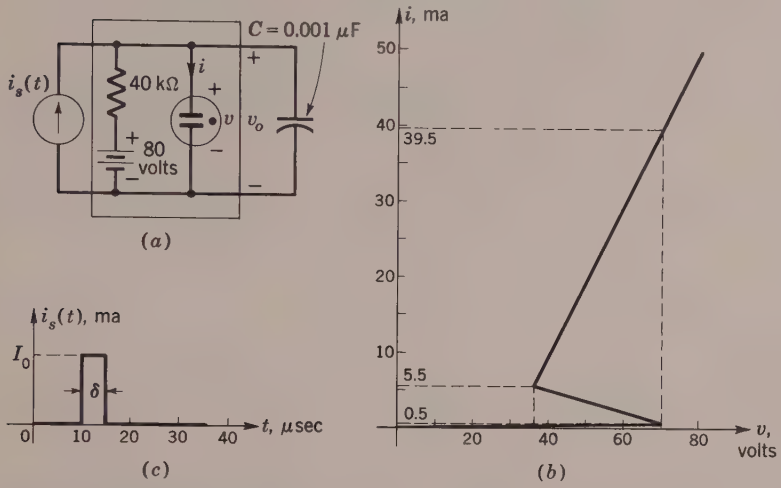


Fig. P16-3.

16-4 A neon lamp multivibrator circuit is shown in Fig. P16-4*a*. If the *v-i* curve of the neon lamp is given by Fig. P16-4*b* and if the trigger waveform $i_s(t)$ is shown in Fig. P16-4*c*, find the output voltage $v_o(t)$ for the following cases:

(a) $\delta = 14\ \mu\text{sec}, I_0 = 0.4$ ma.
(b) $\delta = 14\ \mu\text{sec}, I_0 = 3.0$ ma.
(c) $\delta = 16\ \mu\text{sec}, I_0 = 3.0$ ma.
(d) $\delta = 46\ \mu\text{sec}, I_0 = 3.0$ ma.



(a) (c) (b)

Fig. P16-4.

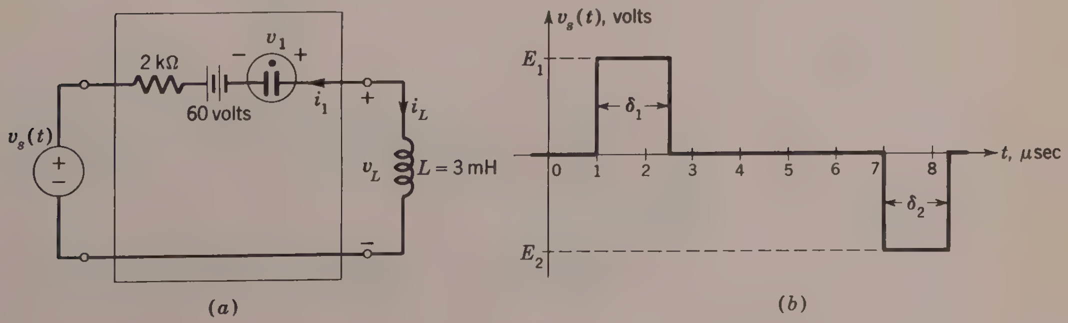
16-5 The triggering signal $v_s(t)$ for the bistable circuit shown in Fig. P16-5*a* is given in terms of the parameters δ_1, δ_2, E_1, and E_2 as shown in Fig. P16-5*b*. The *v-i* curve of the neon lamp is shown in Fig. P16-4*b*.

(a) Find the minimum value of E_1 in order to trigger the circuit from its high-current equilibrium state to its low-current equilibrium state.

(b) If $E_1 = 20$ volts, find the minimum value of δ_1 to ensure a successful triggering operation.

(c) If $E_1 = 20$ volts and $\delta_1 = 1.7\ \mu$sec, find the waveform of $i_C(t)$ for $t \leq 5\ \mu$sec.

(d) Assume that the circuit is in its low-current equilibrium state at $t = 7\ \mu$sec when the negative trigger pulse shown in Fig. P16-5*b* is applied. (1) Find the minimum value of E_2 to ensure a successful triggering operation. (2) If $E_2 = -25$ volts, find the minimum value of δ_2. (3) If $E_2 = -25$ volts and $\delta = 1.4\ \mu$sec, find $i_L(t)$ for $t \geq 7\ \mu$sec.



(a) (b)

Fig. P16-5.

16-6 Most commercial square-wave voltage generators are designed to generate abrupt pulses and then clip them to produce a square wave. It is possible to eliminate the clipper by using the circuit shown in Fig. P16-6.

(a) Explain how this circuit works.

(b) Derive the design equations for determining the values of R, L, and the zener breakdown voltage E_z in order to generate a square voltage wave having a peak-to-peak amplitude of $2E$ volts and a frequency of f kHz.

(c) What are the advantages and possible disadvantages of this circuit?

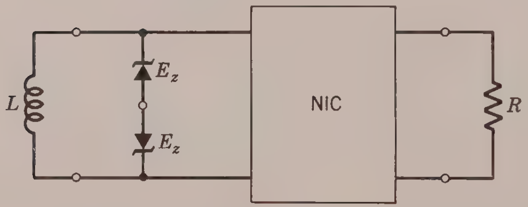


Fig. P16-6.

16-7 One method of generating a square-wave current generator is to make use of two identical constant-current diodes in the circuit shown in Fig. P16-7.

(a) Explain how this circuit works.

(b) Derive the design equations for determining the values of R, C, and the constant current I_0 in order to generate a square current wave having a peak-to-peak amplitude of $2I_0$ ma and a frequency of f kHz.

(c) What are the advantages and possible disadvantages of this circuit?

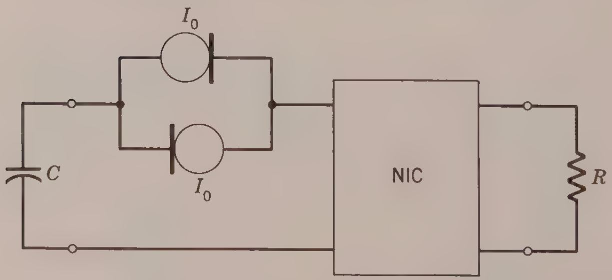


Fig. P16-7.

16-8 The bistable tunnel-diode multivibrator circuit shown in Fig. P16-8*a* is a basic component of many modern digital computers. This circuit can be triggered to furnish two stable values of output voltage $v(t)$. The v-i curves of the tunnel diodes are shown in Fig. P16-8*b* and *c*.

(a) Sketch the waveform of the triggering signal $i_s(t)$ in order to trigger from the low-voltage state to the high-voltage state.

Specify the minimum pulse amplitude and the minimum pulse width corresponding to a pulse amplitude of 10 ma.

(b) Repeat (*a*) in order to trigger from the high-voltage state to the low-voltage state.

(c) Sketch the output voltage waveform corresponding to the triggering pulse shown in Fig. P16-8*d*.

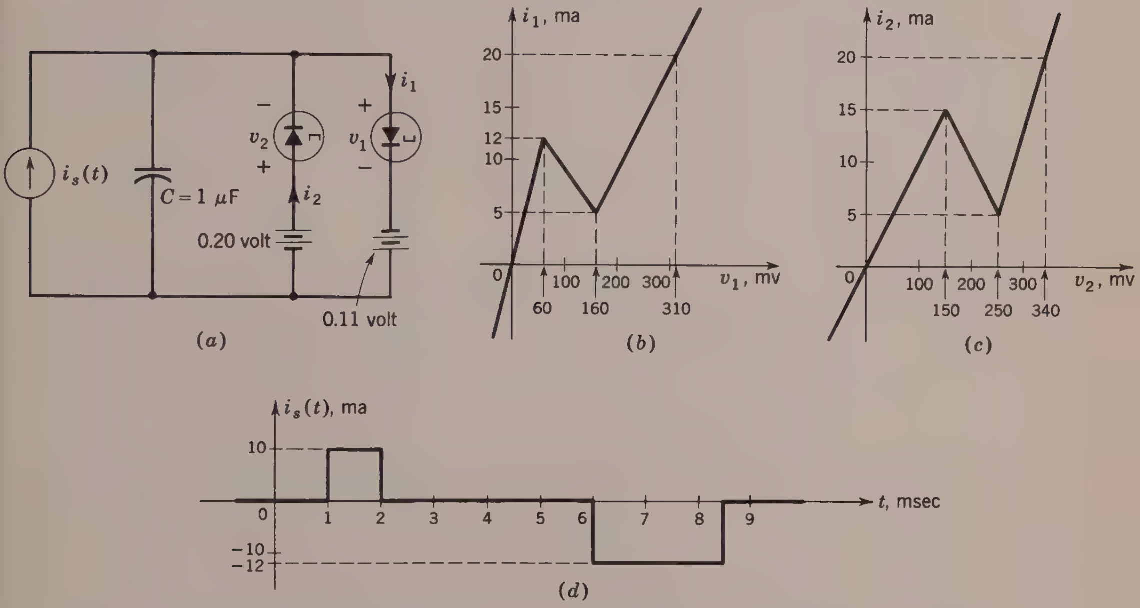


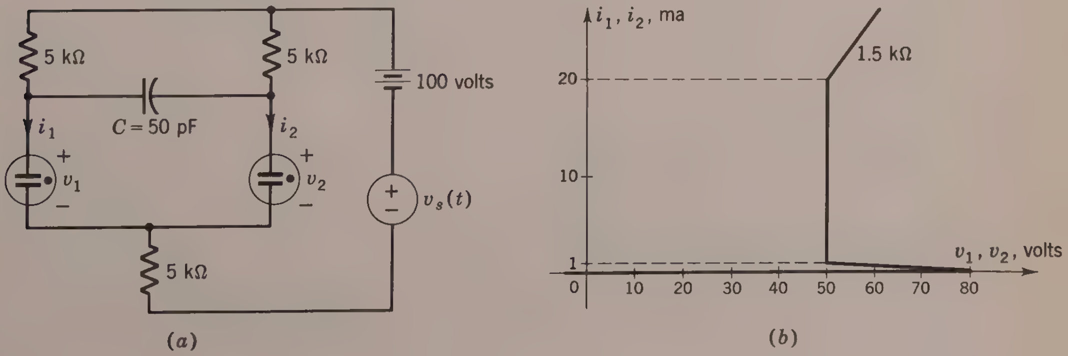
Fig. P16-8.

16-9 Consider the bistable multivibrator circuit shown in Fig. P16-9*a*. The two identical neon bulbs are characterized by the *v-i* curve shown in Fig. P16-9*b*.

(a) Find the voltages $v_1(t)$ and $v_2(t)$ when the circuit is in a stable equilibrium state.

(b) If $v_s(t)$ is a rectangular triggering pulse, find the minimum pulse amplitude for successful triggering.

Fig. P16-9.



16-10 Consider the two basic circuit configurations shown in Fig. 16-19*c* and *d* in the text. Both circuits are made up of a nonlinear resistor (which may be either voltage-controlled or current-controlled), a *linear* biasing circuit, and an energy-storage element.

(a) Show that because the load line is linear, any dynamic route (corresponding to any initial condition) of these circuits can be found by obtaining the composite DP plot of the biasing circuit with portions of the v_d-i_d curve of the nonlinear resistor representing the corresponding locus of operating points.

(b) Using the method described in (*a*), find the current $i_d(t)$ of the circuit shown in Fig. P16-2 with $E = 350$ mv, $R = 35\ \Omega$, $L = 1$ H, and $i_d(0) = 0$.

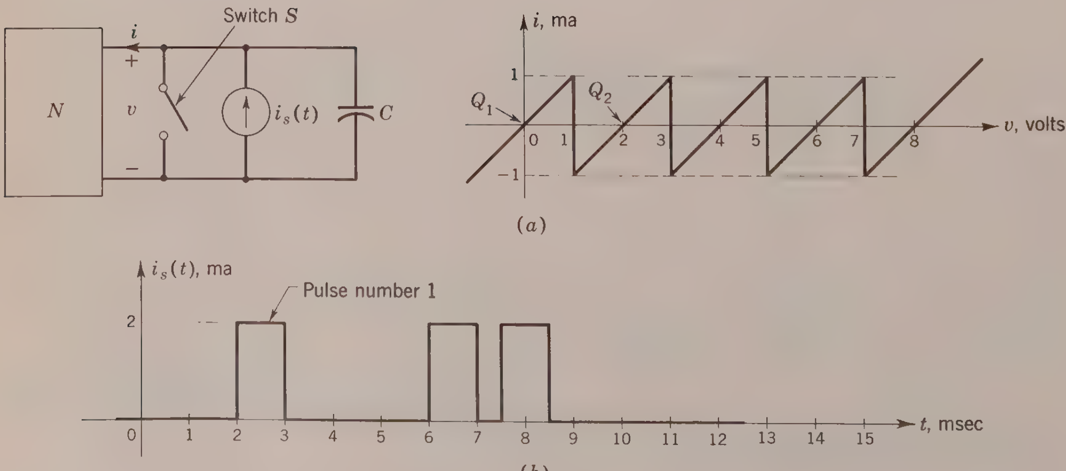
(c) Check the solution in (*b*) by using the method described in the text. Compare the relative advantages of these two methods.

16-11 The circuit shown in Fig. P16-11*a* can be used to determine the number of current pulses (up to a maximum of four) generated by the random pulse current source $i_s(t)$ in a selected 10-msec time interval. The 10-msec interval is selected by opening switch S for 10 msec. Assume that S opens at $t = 0$ and closes at $t = 10$ msec. Further assume that during this 10-msec time interval, the current source $i_s(t)$ generates three 1-msec pulses as shown in Fig. P16-11*b*.

(a) Determine the value of C so that pulse 1 switches the circuit from Q_1 to Q_2 in exactly 1 msec. Sketch the current waveform during the switching cycle.

(b) Find the voltage across the capacitor at $t = 9.99$ msec.

Fig. P16-11.



(*a*)

(*b*)

17 SYNTHESIS OF FIRST-ORDER TIME-BASE GENERATORS: THE BLACK-BOX APPROACH

17-1 WHAT IS A TIME-BASE GENERATOR?

Any network designed specifically to provide a linear time scale is called a time-base generator. This class of networks is indispensable in the design of many electronic systems involving time measurements. Some typical examples are radar, television, cathode-ray oscilloscopes, and timing circuits. As will be obvious soon, either the output voltage or the output current waveform of a time-base generator must be a linear function of time over some specified time interval.

To show the necessity for a linear waveform, let us design a horizontal deflection system for an oscilloscope to sweep the electron beam left to right at a uniform velocity. For convenience, let us assume that the horizontal deflection plates are connected with a battery in series with the output terminals of a time-base generator $v_x(t)$ as shown in Fig. 17-1*a*. The battery voltage E_B is chosen so that when $v_x(t) = 0$, the electron beam will be deflected to the leftmost position A. As $v_x(t)$ increases positively, the electron beam will be swept toward the rightmost position B as shown in Fig. 17-1*b*. Clearly, if $v_x(t)$ is assumed to be a linear waveform as shown in Fig. 17-1*c*, then by projecting the corresponding points as shown in Fig. 17-1*d*, it is clear that the time of occurrence of the electron beam at each point along the horizontal line A-B is uniquely determined. Since $v_x(t)$ is linear, the beam velocity is uniform, and the horizontal axis A-B can be calibrated with a linear time scale. This means that if we apply a voltage signal $v_y(t)$ across the vertical deflection plates, the electron beam will trace out that portion of $v_y(t)$ which occurs during the same time interval it takes the electron beam to sweep from point A to point B.

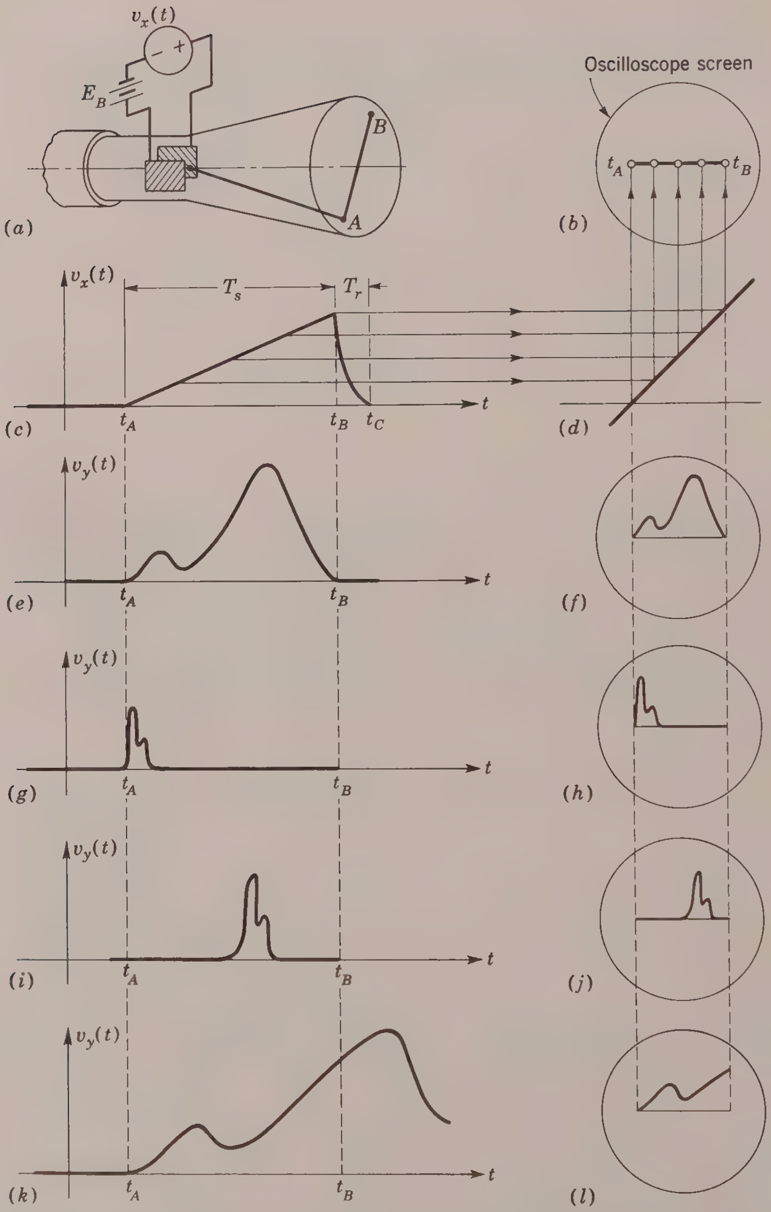


Fig. 17-1. The principle involved in the display of an input voltage signal on the oscilloscope screen requires the application of a linear time-base waveform across the vertical deflection plates of an oscilloscope.

When the beam reaches point B at t_B, it is necessary to make $v_x(t_B) = 0$ in order for the beam to return to point A. However, in view of the principle of conservation of energy, it is impossible for a physical network to exhibit an instantaneous jump. Therefore, the beam will return to point A at some time $t_C > t_B$. It is convenient to define

$$T_s \equiv t_B - t_A \tag{17-1}$$

and

$$T_r \equiv t_C - t_B \tag{17-2}$$

where T_s is called the *sweep time* and T_r is called the *retrace time.*[1] Ideally, we would like the retrace time to be as short as possible.

Let us now investigate the effects of the sweep time T_s on the displayed waveform. Consider first the case where the vertical signal $v_y(t)$ consists of a single pulse. If the pulse width is approximately equal to the sweep time T_s as shown in Fig. 17-1*e*, and if the pulse starts at $t = t_A$, then the entire waveform will be swept and displayed as shown in Fig. 17-1*f*. If the pulse width is much shorter than T_s as shown in Fig. 17-1*g*, the complete waveform will still be displayed as shown in Fig. 17-1*h*. However, the displayed waveform will be so "compressed" that the details cannot be seen. Figure 17-1*i* shows the same input waveform starting at some time $t_A < t < t_B$, and the same compressed waveform is displayed as shown in Fig. 17-1*j*. If the pulse width is much longer than the sweep time as shown in Fig. 17-1*k*, then only a portion of the input signal is displayed as shown in Fig. 17-1*l*. It should be emphasized that since these input signals consist of a single pulse, the corresponding waveform will be displayed only once.[2]

Consider next the case where the input signal $v_y(t)$ is periodic. Suppose it is desired to display the sinusoidal signal $v_y(t)$ shown in Fig. 17-2*a*. Since this signal is periodic, we can obtain a *stationary* display on the oscilloscope screen if the output voltage of the time-base generator consists of a periodic sawtooth waveform as shown in Fig. 17-2*b* to *d*. Since this waveform is similar to the output waveform of an astable (free-running) multivibrator, the network for generating this is called a *free-running time-base generator.* The sweep time T_s of the sawtooth waveform shown in Fig. 17-2*b* is approximately equal to three times the period of $v_y(t)$, and hence approximately three cycles of the sinusoidal signal will be displayed. Similarly, the sweep time of the sawtooth waveforms shown in Fig. 17-2*c* and *d* will result in a stationary display of approximately two cycles and one cycle of the sinusoidal signal, respectively. In all cases observe that the sweep time T_s and the retrace time T_r were chosen so that the starting point of each new cycle of the sawtooth signal always coincides with the same position

---

[1] The retrace time is also known as the return time, the restoration time, or the flyback time.

[2] If it is desired to have the displayed waveform persist on the oscilloscope screen, a special-purpose oscilloscope called the *memory scope* must be used.

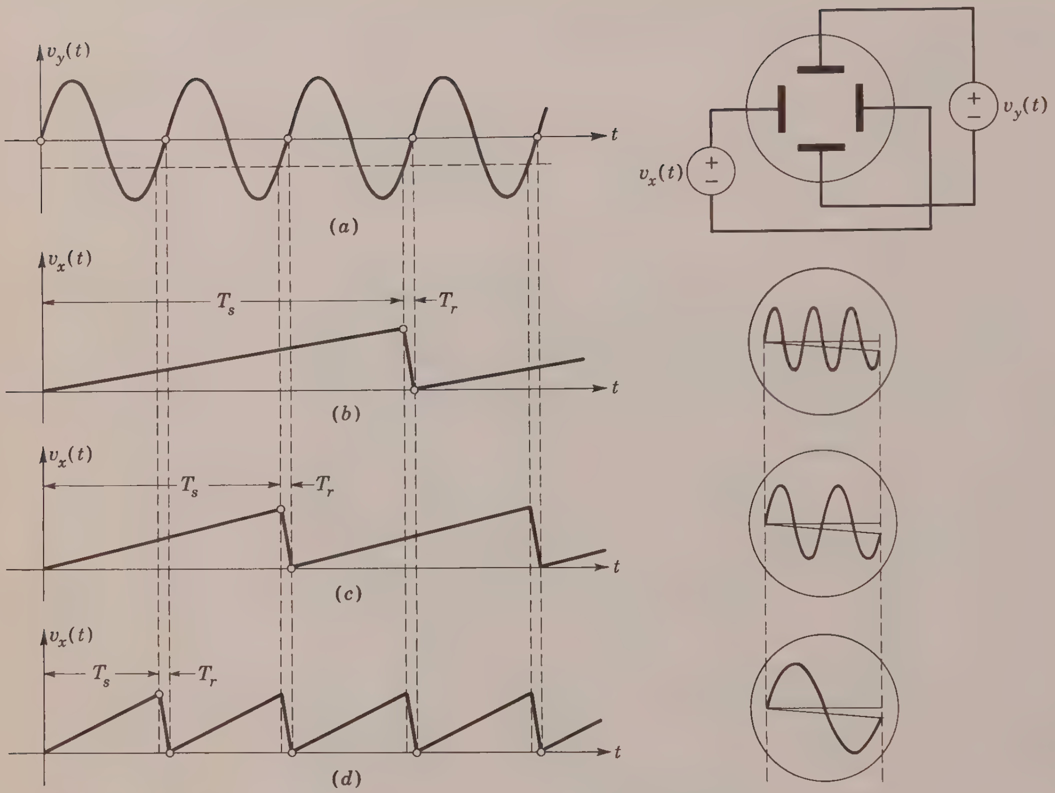


Fig. 17-2. A free-running sawtooth waveform is required to display a stationary periodic signal.

on the sinusoidal waveform. This condition is clearly necessary in order to display a stationary waveform on the oscilloscope screen. Observe that since the retrace time T_r is not zero, the corresponding portion of the sinusoidal signal will be displayed in the return trace as shown in Fig. 17-2*b* to *d*. This return trace can be eliminated by shutting off the electronic beam during the retrace time intervals. This is easily achieved by applying a high negative voltage called the "blanking signal" to the controlled grid of the electron gun.

Any number of complete cycles of the sinusoidal waveform shown in Fig. 17-2*a* can be displayed steadily on the oscilloscope screen by an appropriate choice of the sweep time T_s. For example, if a detailed examination of the displayed waveform is desired, a single cycle can be displayed across the oscilloscope screen as shown in Fig. 17-2*d*. This flexibility cannot be attained if the periodic signal consists of a widely separated pulse train such as the one shown in Fig. 17-3*a*. Observe that in order to obtain a stationary display of this signal by means of a free-running time-base generator, the period of the sawtooth waveform must be at least as large as the period of the pulse train as shown in Fig. 17-3*b*. Since the pulse width δ is extremely short compared with the period T, the displayed waveform will be so compressed that it will appear almost like a vertical line. In order to spread out this pulse, it is necessary that the sweep time T_s be equal to the pulse width. Moreover, in order for the displayed waveform to persist on the oscilloscope screen, it is necessary that whenever a pulse occurs, a linear time-base waveform is generated as shown in Fig. 17-3*c*. This can be done by a triggering circuit, and the resulting network is called a *triggered time-base generator.*

The preceding discussion shows that there are two types of time-base generators, a free-running time-base generator, which generates a periodic sawtooth waveform, and a triggered time-base generator, which generates a linear waveform with a prescribed duration whenever a triggering signal is applied. Although a time-base *voltage* waveform is required in an oscilloscope, a time-base *current* waveform may be required in other systems, such as in the deflection coils of a television receiver.

Based on the preceding discussions, we can now define an ideal free-running time-base generator as a black box whose output voltage or current consists of a periodic sawtooth waveform

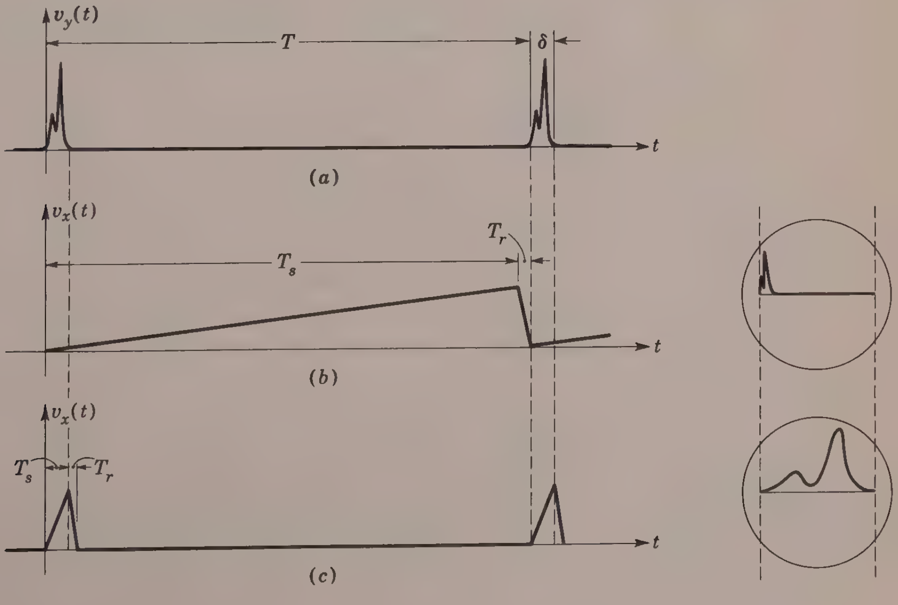


Fig. 17-3. A triggered sawtooth waveform is required to display a widely separated periodic pulse train.

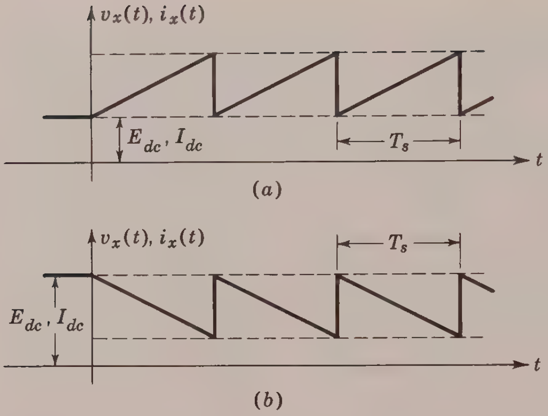


Fig. 17-4. Two common ideal output waveforms of a free-running time-base generator.

with a linear ramp and a zero retrace time. Two common examples of this ideal sawtooth waveform are shown in Fig. 17-4*a* and *b*. Similarly, an ideal triggered time-base generator can be defined as a black box with a triggering source such that a linear output voltage or current waveform is generated each time the switching signal is applied and then drops back to the original value when the switching signal returns to zero. For example, Fig. 17-5 shows a switching signal $v_s(t)$ or $i_s(t)$ with two ideal output waveforms. Observe that for complete generality, the output waveforms of Figs. 17-4 and 17-5 are drawn with a reference voltage E_{dc} or reference current I_{dc}. Observe also that the linear ramp may be either increasing or decreasing with time. Although the ideal

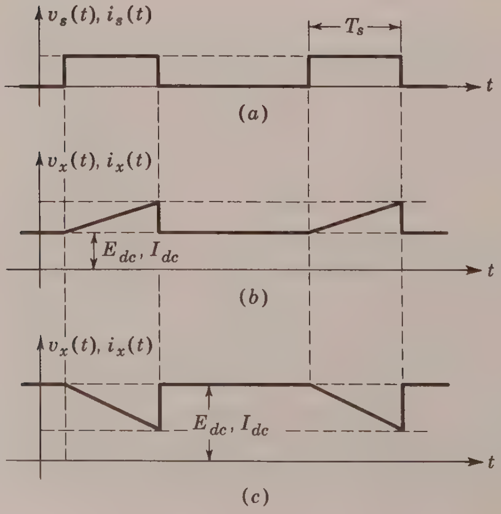


Fig. 17-5. Two common ideal output waveforms of a triggered time-base generator and the associated triggering signal.

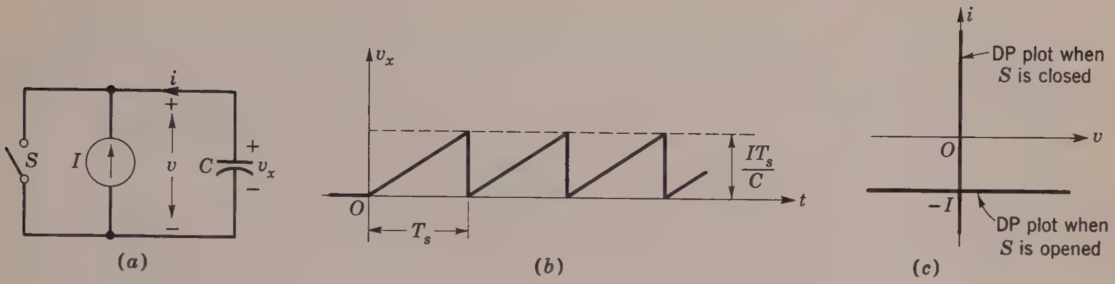


Fig. 17-6. An ideal sawtooth voltage waveform can be generated by a current source in parallel with a switch and a capacitor.

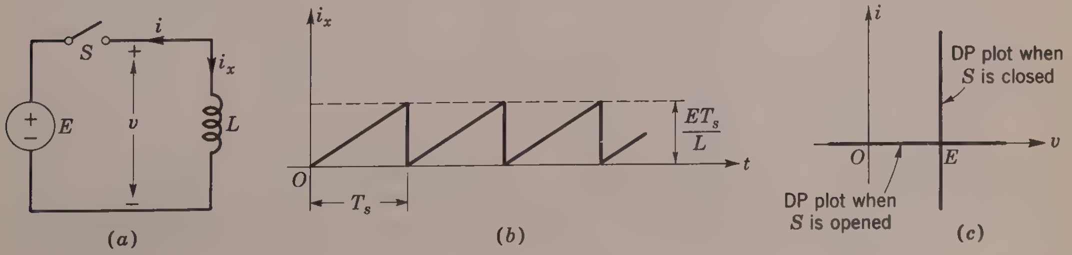
waveforms shown in Figs. 17-4 and 17-5 are not physically realizable, they are useful because they provide a model for the design of practical circuits.

Most existing time-base generators can be considered as first-order nonlinear networks. It is not our objective to study the countless variations of existing circuits because this can at best be a systematic collection of "recipes," a task which we abhor. Instead we shall present only the basic principles underlying the synthesis of time-base generators. In particular, only first-order time-base generators will be considered in this chapter.

17-2 SYNTHESIS OF FREE-RUNNING TIME-BASE GENERATORS

Let us consider first how the ideal waveforms of Fig. 17-4 can be synthesized using ideal elements. We have shown in Chap. 14 that a linear voltage waveform can be generated by connecting a current source across a capacitor. We have also shown that a linear current waveform can be generated by connecting a voltage source across an inductor. Hence, if we connect a switch S across the capacitor C as shown in Fig. 17-6*a*, then by closing and opening the switch S periodically, the ideal voltage waveform shown in Fig. 17-6*b* is obtained. The ideal DP plot seen by the capacitor in this case is shown in Fig. 17-6*c*. Similarly, the series switching circuit shown in Fig. 17-7*a* will have an output sawtooth current as shown in Fig. 17-7*b* and an ideal DP plot as shown in Fig. 17-7*c*. Since

Fig. 17-7. An ideal sawtooth current waveform can be generated by a voltage source in series with a switch and an inductor.



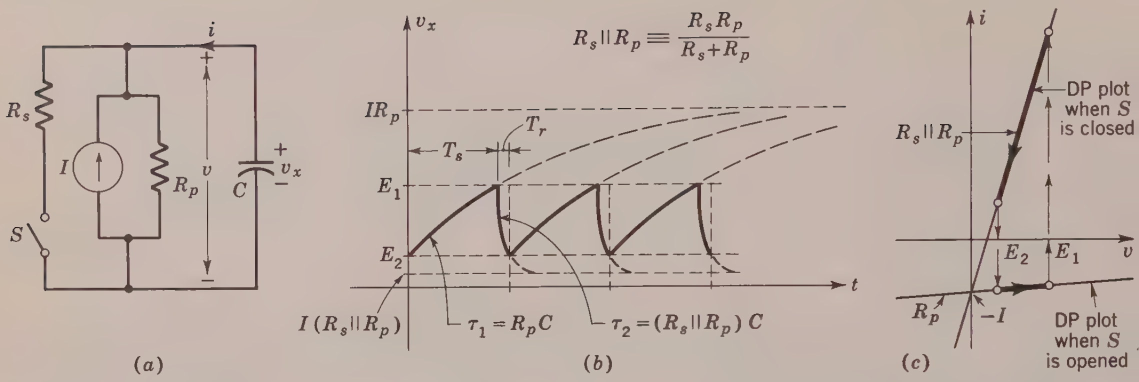


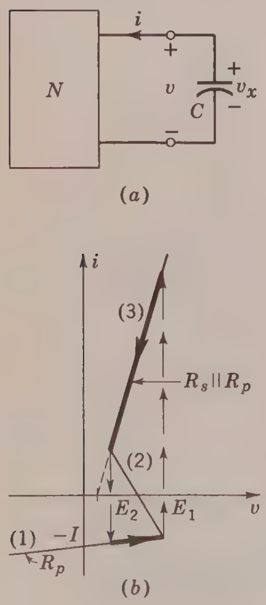
Fig. 17-8. The output voltage waveform and DP plot of a realistic time-base generator.

the current time-base generator is simply the dual of the voltage time-base generator, it suffices to consider only one case from here on.

Since any physical switch has a small but nonzero series resistance R_s and any physical current source has a large but finite parallel resistance R_p, the ideal circuit of Fig. 17-6*a* must be replaced by the more realistic model shown in Fig. 17-8*a*. The presence of the resistances R_s and R_p will prevent any abrupt change in the capacitor voltage. In fact, in order to allow time for the capacitor to discharge quickly during the return trace, the switch *S* must remain closed for a short but nonzero time interval. If we assume that *S* is open until the capacitor charges up to some specified voltage $v_x = E_1$ volts and is then closed until the capacitor discharges down to some specified voltage $v_x = E_2$ volts, then the waveform $v_x(t)$ will be as shown in Fig. 17-8*b*. The associated DP plot and dynamic route are shown in Fig. 17-8*c*.

Since $v_x(t)$ shown in Fig. 17-8*b* is periodic and resembles the output waveform of an astable multivibrator, we can eliminate the switch *S* and the current source *I* and choose, instead, a resistive black box *N* (Fig. 17-9*a*) with the DP plot shown in Fig. 17-9*b*. Since the dynamic route remains unchanged, the same output voltage is obtained. Clearly, it is desirable that segment 1 be as horizontal as possible and segment 3 be as vertical as possible. The main task in designing a free-running time-base generator, therefore, consists of realizing a resistive network with the desired DP plot. Let us consider two simple examples.

Fig. 17-9. A current-controlled DP plot can be chosen in place of the resistive network across the capacitor.



EXAMPLES

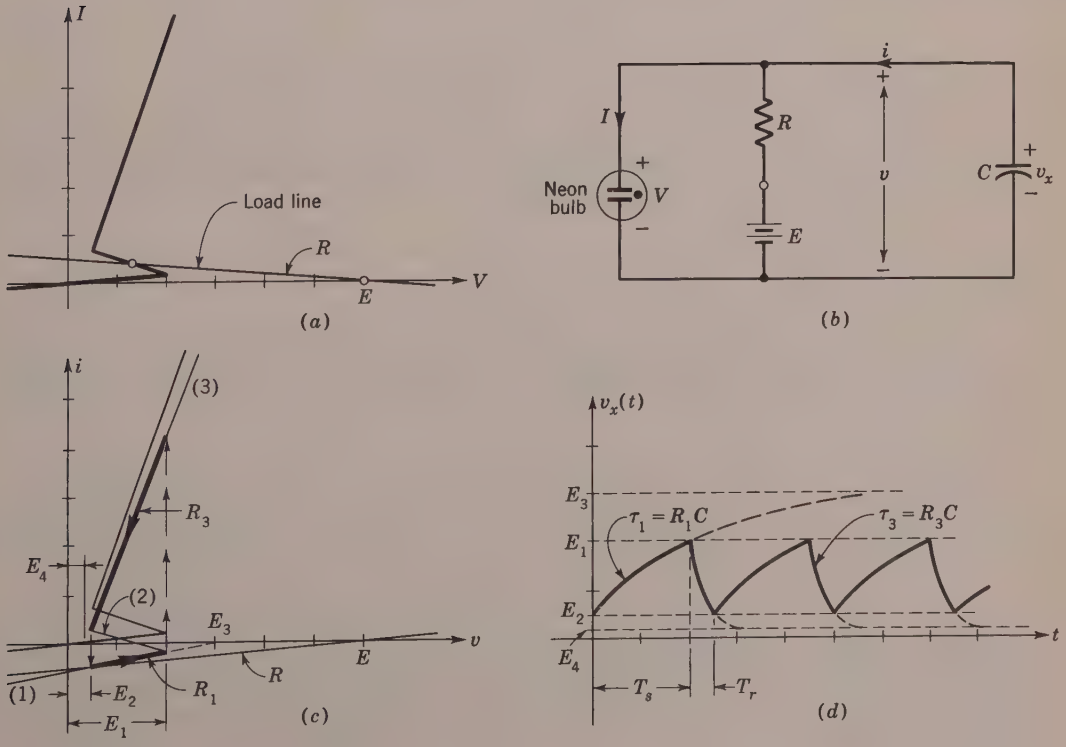
1. *Neon-bulb Time-base Generator.* There are many practical nonlinear resistors and resistive black boxes which exhibit a DP

plot suitable for designing time-base generators. For simplicity, we shall choose a neon bulb whose *v-i* curve is shown in Fig. 17-10*a*. Since a free-running time-base generator is simply an astable multivibrator with a specific DP plot, we can apply the techniques of Chap. 16 to bias the neon bulb for astable operation. The resulting circuit is shown in Fig. 17-10*b* where the values of *R* and *E* are chosen so that the load line will intersect the negative-resistance segment as shown in Fig. 17-10*a*. The composite DP plot is obtained in Fig. 17-10*c* together with the associated dynamic route. Applying the methods of Chap. 15, the output voltage $v_x(t)$ is easily obtained and is shown in Fig. 17-10*d*. The sweep time T_s and retrace time T_r are given, respectively, by

$$T_s = \tau_1 \ln \frac{E_2 - E_3}{E_1 - E_3} \tag{17-3}$$

$$T_r = \tau_3 \ln \frac{E_1 - E_4}{E_2 - E_4} \tag{17-4}$$

Fig. 17-10. The synthesis of a neon-bulb time-base generator.



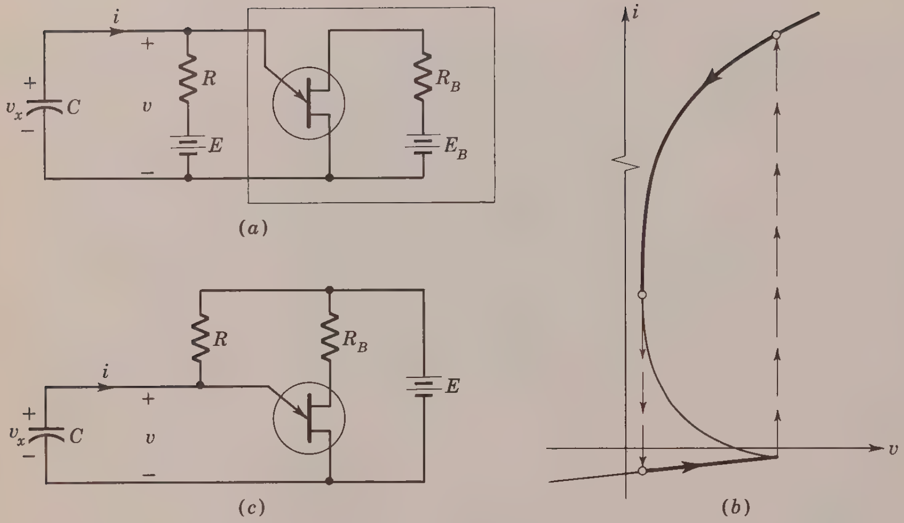


Fig. 17-11. A unijunction transistor time-base generator.

where $\tau_1 = R_1C$ and $\tau_3 = R_3C$. The frequency f_x of this periodic waveform is, therefore, given by

$$f_x = \frac{1}{R_1C \ln \dfrac{E_2 - E_3}{E_1 - E_3} + R_3C \ln \dfrac{E_1 - E_4}{E_2 - E_4}} \tag{17-5}$$

Observe that in order to approach the ideal sawtooth waveform, it is necessary that $\tau_1 \gg \tau_3$. Under this condition Eq. (17-5) can be approximated by

$$f_x \cong \frac{1}{R_1C \ln \dfrac{E_2 - E_3}{E_1 - E_3}} \tag{17-6}$$

In any case the frequency f_x depends on the capacitance C, the supply voltage E,[†] and the slopes of the DP plot. Obviously, the simplest way to change the frequency is to vary the value of C.

2. *Unijunction Transistor Time-base Generator.* Consider the first-order unijunction transistor circuit shown in Fig. 17-11*a*. The DP plot seen by the capacitor was obtained in Fig. 6-40 of Chap. 6 and is redrawn in Fig. 17-11*b*. Notice that the vertical scale is compressed in order to show the complete dynamic route. Since this dynamic route resembles the one shown in Fig. 17-10*c*, the unijunction transistor circuit in Fig. 17-11*a* is in fact a voltage time-base generator. In practice we can save one battery by choosing $E = E_B$ and applying the

[†] Changing the value of E has the effect of translating the composite DP plot vertically and hence changing the values of E_3 and E_4.

v-shift theorem to obtain the equivalent time-base generator circuit shown in Fig. 17-11c.

Exercise 1: Describe the detailed procedure for designing a current time-base generator.

Exercise 2: Using a type 2NXX21 unijunction transistor, design a 10-kHz voltage time-base generator.

17-3 SYNCHRONIZATION OF FREE-RUNNING TIME-BASE GENERATORS

As we have shown earlier, a free-running time-base generator is required to display a stationary periodic signal. In order for the displayed waveform to remain stationary on the oscilloscope screen, it is necessary that the input signal and the sawtooth waveform be "synchronized" at all times. By this we mean that the beginning of each sweep must occur at the same position on the input waveform. For example, consider the periodic input signal $v_y(t)$ shown in Fig. 17-12a and the sawtooth waveform $v_x(t)$ shown in Fig. 17-12b. These two waveforms are synchronized because the

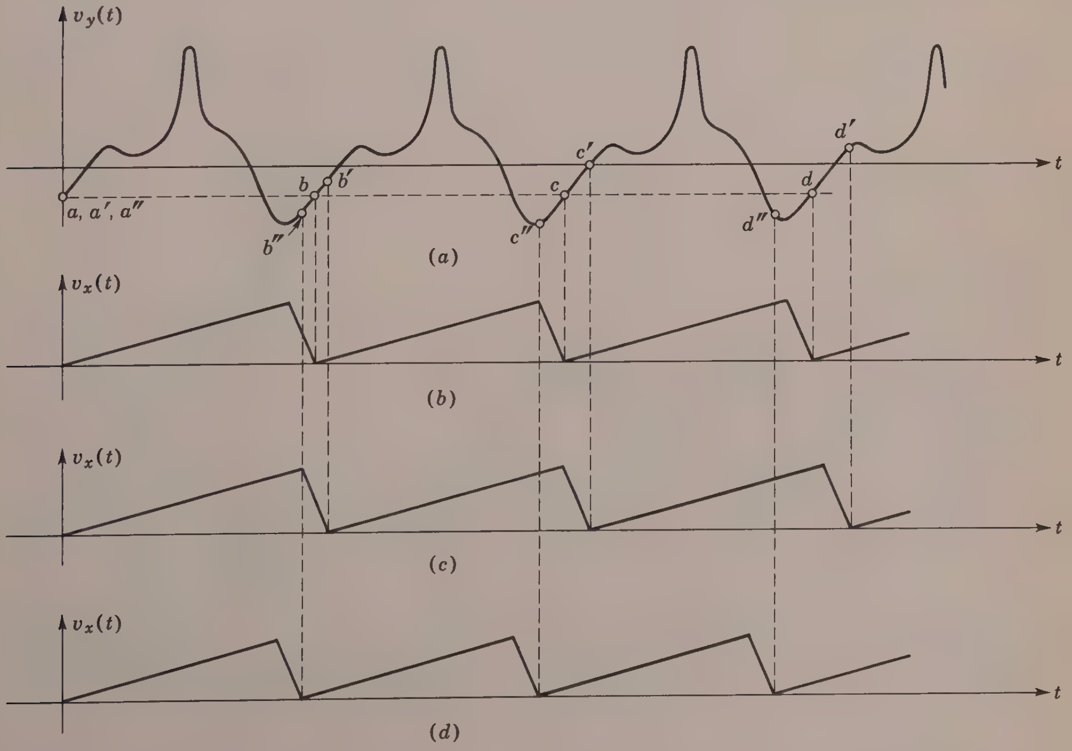


Fig. 17-12. An example showing the need for synchronization between the time-base generator and the input signal.

starting point of each sweep occurs at exactly the same position ($a, b, c, d, \ldots$) along the input signal. Consider now what happens if the frequency of the sawtooth waveform changes by a small increment. If the frequency decreases slightly as shown in Fig. 17-12*c*, the waveforms are no longer synchronized because the starting points $a', b', c', d', \ldots$ no longer coincide with one another. The effect of this is that the displayed waveform will drift slowly to the left of the oscilloscope screen. If the frequency increases slightly as shown in Fig. 17-12*d*, the waveforms are also not synchronized because the starting points $a'', b'', c'', d'', \ldots$ are different. In this case the displayed waveform will tend to drift slowly to the right. From these observations we can conclude that in order for synchronization to be possible, the frequency f_x of the time-base generator must be an integral submultiple of the input signal frequency f_y; namely,

$$f_x = \frac{1}{n} f_y \tag{17-7}$$

where n is any integer equal to or greater than 1.

Since the frequency f_x depends on such parameters as the supply voltage and the slopes of the DP plot, it is clear that even if Eq. (17-7) was initially satisfied, the slightest perturbation due to noise or other extraneous factors will immediately result in a loss of synchronization. There are various methods that can be used to force the time-base frequency f_x to "synchronize" or "lock" with the input signal frequency f_y. Since the basic principles are the same in all cases, we shall consider only one method, the *pulse synchronization method.*

In the pulse synchronization method the time-base generator is synchronized with a pulse train $v_s(t)$ having the same frequency f_y as the input signal. This synchronization signal $v_s(t)$ is always derived from the input signal $v_y(t)$ by some standard waveform operation in order to ensure that the frequency is exactly f_y. For example, $v_s(t)$ can be obtained from the output of a monostable multivibrator which is being triggered by the input signal $v_y(t)$. Ideally, each synchronization pulse should have a very fast rise time. Let us consider the mechanism behind the synchronization phenomenon.

The DP plot and the output waveform of a typical time-base generator are shown in Fig. 17-13*a* and *b*. The dynamic route jumps from segment 1 to segment 3 whenever the breakpoint $B(v = E_1)$ is reached. This is indicated in Fig. 17-13*b* by the horizontal switching threshold $v = E_1$. If E_1 either increases or decreases, this switching threshold will move either up or down. Accordingly, the frequency f_x of $v(t)$ will either decrease or increase. In practice, because of the presence of noise or other extraneous factors such as changes in temperature and environment, the DP plot of Fig. 17-13*a* will vary randomly. This variation can be represented by a tolerance band around the DP plot as shown in Fig. 17-13*c*. In other words, because of the uncertainties involved, we can state only that the DP plot at any time will lie within the tolerance region. If we denote the maximum variation in v by

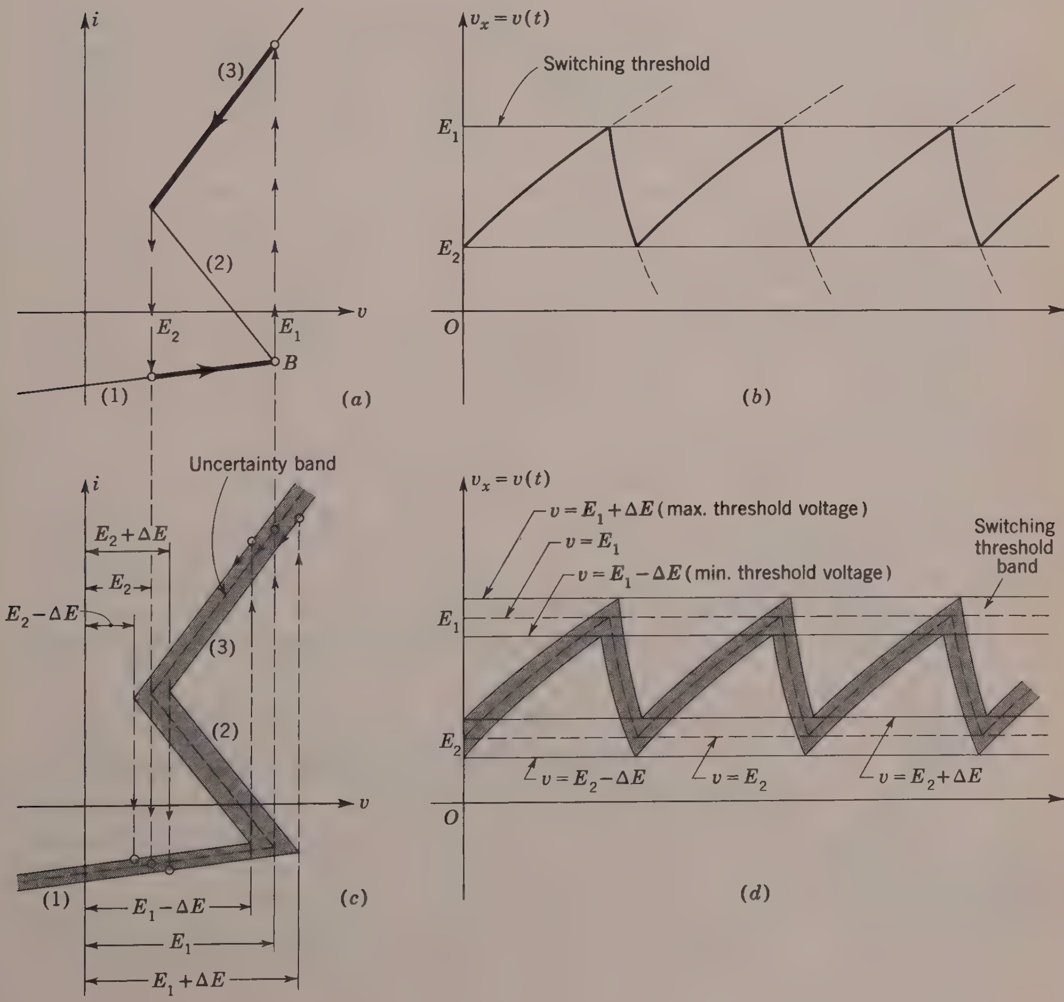


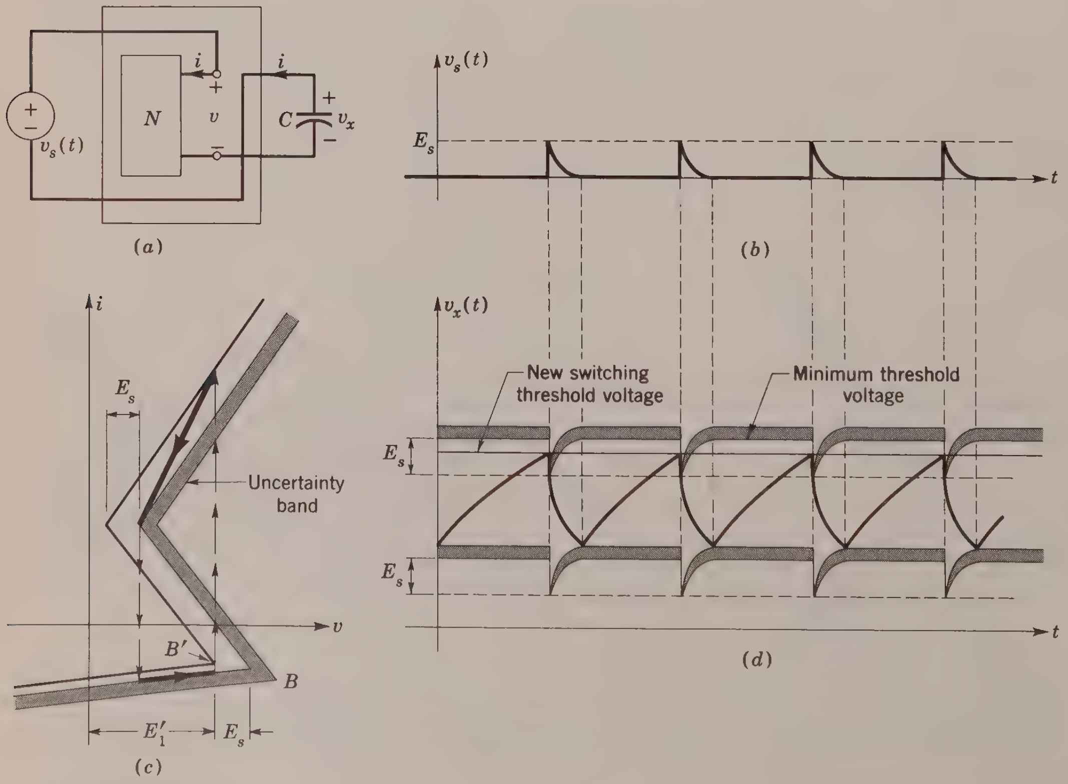
Fig. 17-13. The frequency of any physical time-base generator will vary whenever the switching threshold changes.

$\pm\Delta E$, then the horizontal switching threshold becomes a threshold band as shown in Fig. 17-13*d*.

Let us observe what happens if we connect a synchronization voltage source $v_s(t)$ in series with the capacitor as shown in Fig. 17-14*a*. If $v_s(t)$ is as shown in Fig. 17-14*b*, then the DP plot will be shifted to the left whenever a pulse appears, and the threshold voltage will decrease by an amount equal to the pulse height E_s of $v_s(t)$. The switching threshold will, therefore, change as shown in Fig. 17-14*d* whenever a pulse occurs. Observe that because we have hastened the switching to occur prematurely, i.e., before the dynamic route reaches the uncertainty band, the frequency of $v_x(t)$ in Fig. 17-14*d* will be precisely equal to that of $v_s(t)$. Therefore, the time-base generator is synchronized with the input signal, and one cycle of $v_y(t)$ will remain stationary on the oscilloscope screen.

We shall now show that a minimum pulse height and a minimum pulse frequency (relative to the unsynchronized frequency f_x) is required to ensure synchronization. Consider first the case

Fig. 17-14. The mechanism for synchronization consists of initiating the vertical jump before the minimum threshold voltage is reached.



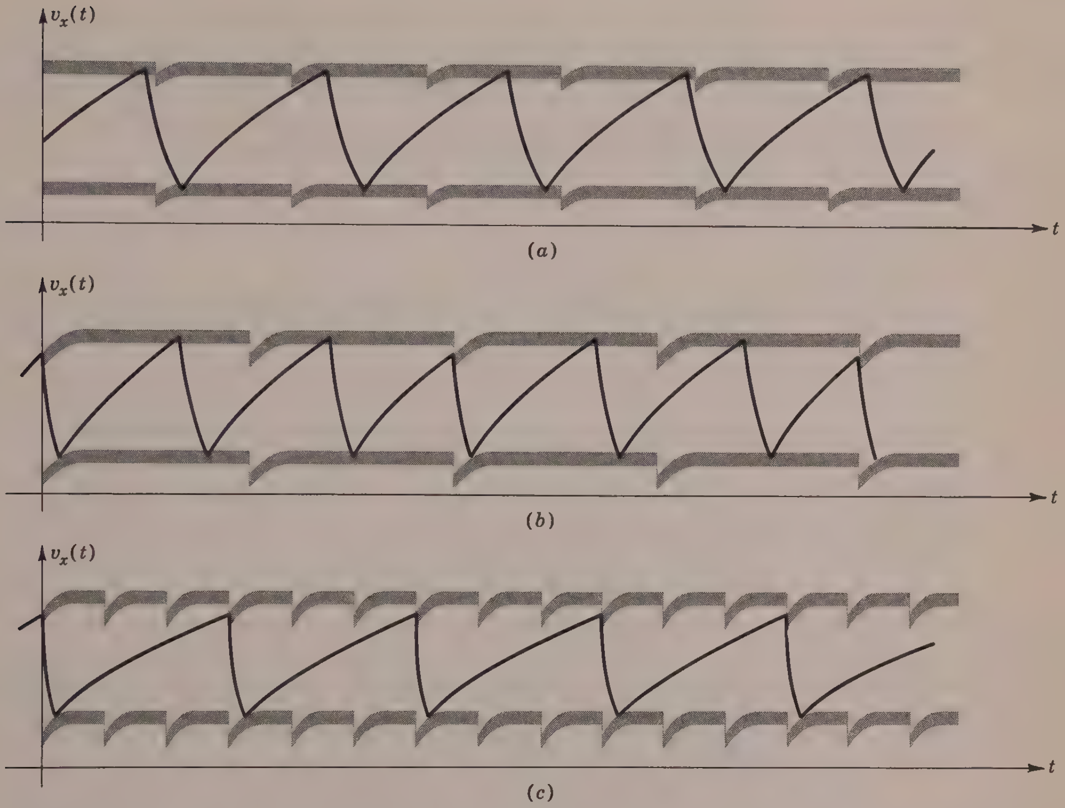


Fig. 17-15. Waveform relationships showing that synchronization will not occur whenever (*a*) the pulse height or (*b*) the pulse frequency is too small. Synchronization can be attained, however, with (*c*) a higher-frequency synchronization signal.

when the pulse height is relatively small as shown in Fig. 17-15*a*. It is easily seen that this pulse height is too small to cause the switching to occur prematurely, and hence no synchronization is possible in this case. For synchronization to occur, the minimum pulse height must be chosen [depending on the frequency of $v_s(t)$] so that the breakpoint B of the DP plot of Fig. 17-14*c* is shifted to the left of the value assumed by $v_x(t)$ at the time when a pulse occurs.

When the pulse frequency is smaller than the unsynchronized frequency f_x of the time-base generator as shown in Fig. 17-15*b*, again no synchronization is possible.[1] It is easy to see that in order for synchronization to occur, the frequency of $v_s(t)$ must be greater than the highest possible unsynchronized frequency f_x of the time-base generator. An example illustrating this condition is shown in Fig. 17-15*c*. It is easy to see that, provided the pulse height is large enough, the sawtooth waveform $v_x(t)$ will synchronize with a subharmonic frequency of $v_s(t)$. In this case, more than one stationary cycle of the input signal can be displayed on the oscilloscope screen. This phenomenon is useful in many other applications. For example, if we consider $v_s(t)$ as the input signal and $v_x(t)$ as the output signal, then the output frequency f_x will be equal to a subharmonic of the input frequency f_s; namely,

---

[1] Observe that in this case a partial synchronization occurs every few cycles. However, this phenomenon does not have any practical value at present.

$$f_x = \frac{1}{n} f_s \tag{17-8}$$

This property is the basic operating principle of most frequency dividers that exist today. Indeed, by connecting an appropriate number of these dividers in cascade, it is possible to divide accurately an extremely high-frequency pulse train down to any desired output frequency. This property has been used in the design of an atomic clock. For example, certain atomic beams (such as that of cesium) are known to resonate at an extremely high and extremely stable frequency. A typical resonant frequency might be 10^{10} cps with a maximum frequency variation of ΔF equal to 10^{-5} cps. If we divide this frequency down to 1,000 cps to operate a synchronous motor, then $n = 10^7$, and the output frequency variation is $\Delta f = 10^{-5}/10^7 = 10^{-12}$ cps. Hence it will take 10^{12} sec for the motor frequency to change by one cycle. In view of this great stability, the motor can be calibrated to serve as an extremely accurate clock.[1]

We have merely scratched the surface of the subject of synchronization. This phenomenon is unique in that it can occur only in nonlinear systems. Although a complete theory of synchronization is not presently available, it is good to know that most practical synchronization mechanisms are similar to the simple case presented in this section.

Exercise 1: Discuss the synchronization mechanism using a periodic rectangular synchronization pulse.

Exercise 2: Discuss the synchronization mechanism using a sinusoidal synchronization signal.

17-4 SYNTHESIS OF TRIGGERED TIME-BASE GENERATORS

We have already shown that in order to obtain a *stationary* display of one pulse of a widely separated pulse train and spread it out across the entire oscilloscope screen, it is necessary to initiate a linear sweep each time a pulse occurs. Ideally, this can be done by opening a normally closed switch S in the circuit of Fig. 17-6*a*

---

[1] The rpm of a synchronous motor is a function of the frequency of the applied voltage.

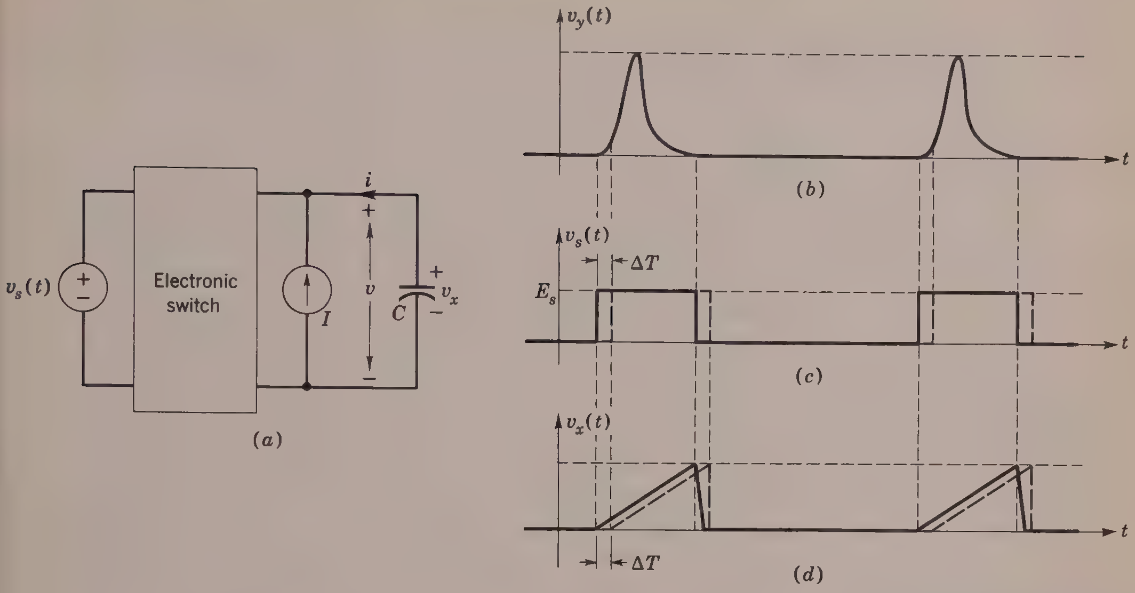


Fig. 17-16. An ideal voltage time-base generator being triggered by an electronic switch.

during the time interval where a pulse is present. The electronic switching circuit shown in Fig. 17-16*a* can be used to open the switch at the start of each pulse and to close the switch upon the termination of the pulse. The switching signal $v_s(t)$ is generally a square wave obtained from the input pulse train. For example, the typical pulse train $v_y(t)$ shown in Fig. 17-16*b* can be used to trigger a monostable multivibrator. The output of the multivibrator can then be clipped to obtain the desired switching signal $v_s(t)$ shown in Fig. 17-16*c*. In this case the electronic switch must be designed so that it will be open whenever $v_s(t) = E_s$ and closed whenever $v_s(t) = 0$. The output waveform $v_x(t)$ of this ideal time-base generator will then be shown in Fig. 17-16*d*.

Several practical limitations must be considered in the design of a time-base generator. First of all, the switching signal $v_s(t)$ will generally be delayed by some short but nonzero time interval ΔT. In this case, $v_s(t)$ and $v_x(t)$ will also be delayed by ΔT as shown by the dotted lines. The effect of this is that a small initial portion of the input waveform $v_y(t)$ will not be displayed. However, by a careful choice of components, it is possible to reduce ΔT to a very small value so that the portion of $v_y(t)$ that is lost will not be noticeable. Another practical limitation is the fact that it is impossible to design an ideal electronic switch or an ideal current source. This means that the output waveform $v_x(t)$ of the time-base generator will not be linear and that the retrace time will not be zero. With these limitations in mind, let us now consider a few sample designs.

EXAMPLES

1. *An FET-triggered Time-base Generator.* From the basic circuit configuration shown in Fig. 17-16*a*, it is clear that a triggered time-base generator can be designed in two steps: (1) choose or design a suitable current source, and (2) design the associated electronic switch.

For step 1 let us refer to Fig. 17-8 and observe that it is not really necessary to have a current source. Indeed, any two-terminal black box whose DP plot contains a horizontal segment which intersects the *i* axis (Fig. 17-17) would do just as well. For example, the DP plot of each of the three black boxes shown in Fig. 17-18*a* to *c* contains a nearly horizontal segment and can, therefore, be used in place of the ideal current source. The value of *R* in Fig. 17-18*a* must be chosen very large in order for the DP plot to approach a horizontal line. In all cases, however, the segments will always have a nonzero slope. It is possible to reduce the slope of these segments by adding compensating circuits, thus improving the linearity of the output waveform. Since this refinement does not involve any new concept, we shall not bother to consider it here.

For step 2 we can apply the principles presented in Chap.

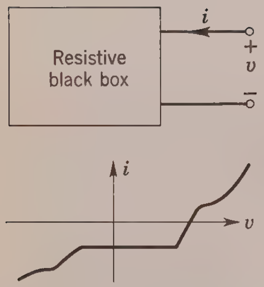
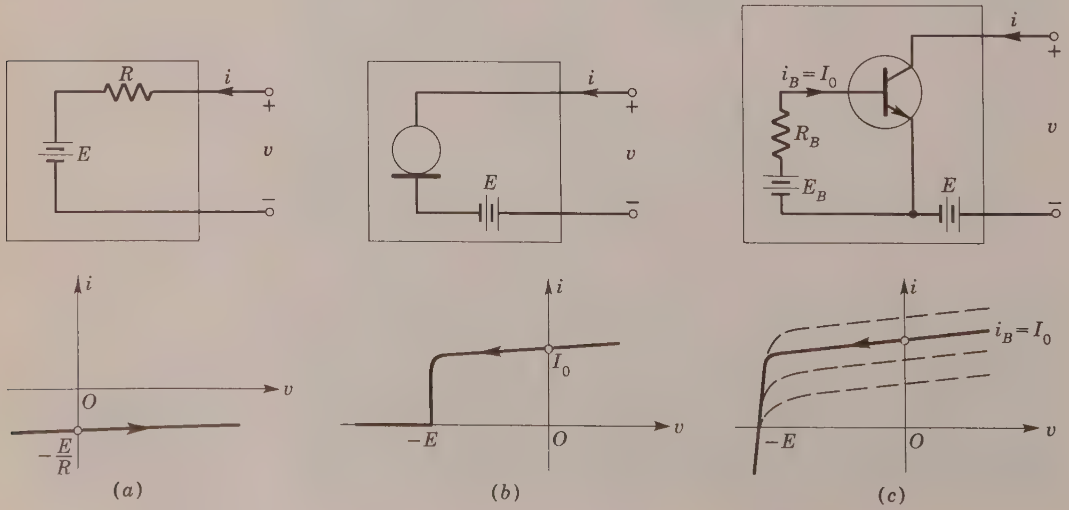


Fig. 17-17. A DP plot with a horizontal segment intersecting the *i* axis can be used in place of the current source.



(*a*) (*b*) (*c*)

Fig. 17-18. Three simple practical circuits that can be used in place of the current source.

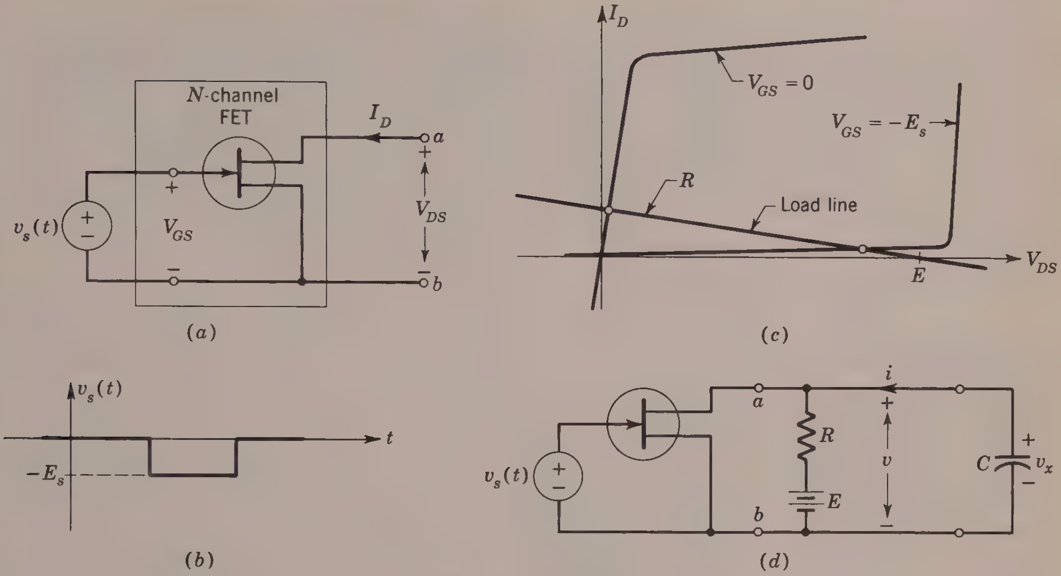
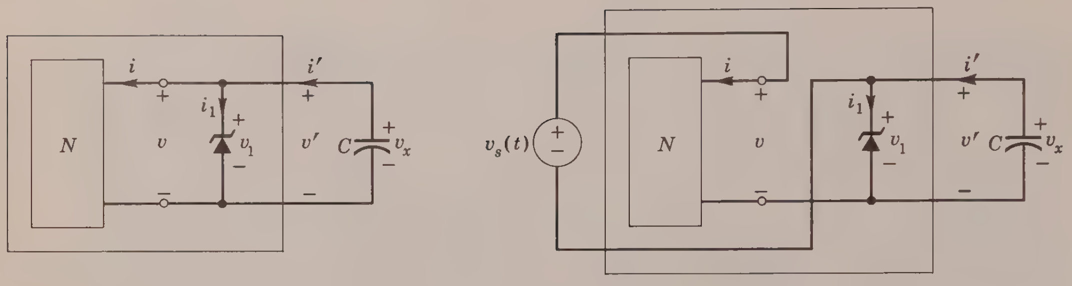


Fig. 17-19. The complete design of a simple FET-switch-triggered time-base generator.

10 to synthesize the appropriate electronic switch. For example, if we choose the *n*-channel FET switch shown in Fig. 17-19*a* and the switching signal $v_s(t)$ shown in Fig. 17-19*b*, then the DP plot across terminals *a-b* will approximate an open circuit when $V_{GS} = -E_s$ and a short circuit when $V_{GS} = 0$. Since this is a *unidirectional* switch, we must choose the circuit of Fig. 17-18*a* to simulate the current source so that the corresponding load line will intersect the horizontal segment when $V_{GS} = -E_s$ and the vertical segment when $V_{GS} = 0$ (Fig. 17-19*c*). The complete triggered time-base generator is shown in Fig. 17-19*d*. Observe that since the current intercept of the load line corresponding to the circuits of Fig. 17-18*b* and *c* is negative, it would be necessary to use a *p*-channel FET switch if these circuits were to be used in place of the ideal current source.

2. *A Self-triggered Time-base Generator.* Another method for designing a triggered time-base generator is to synthesize first a free-running time-base generator and then prevent the switching operation from occurring. A simple method to accomplish this is to connect a zener diode across the resistive black box of a free-running time-base generator as shown in Fig. 17-20*a*. If the original DP plot of the resistive black box is shown in Fig. 17-21*a* and the DP plot of the inverted zener diode is shown in Fig. 17-21*b*, then the composite DP plot is readily ob-



(a) (b)

Fig. 17-20. A free-running time-base generator can be converted into a triggered time-base generator by connecting a switching voltage source in series with the resistive black box and a zener diode in parallel with the capacitor.

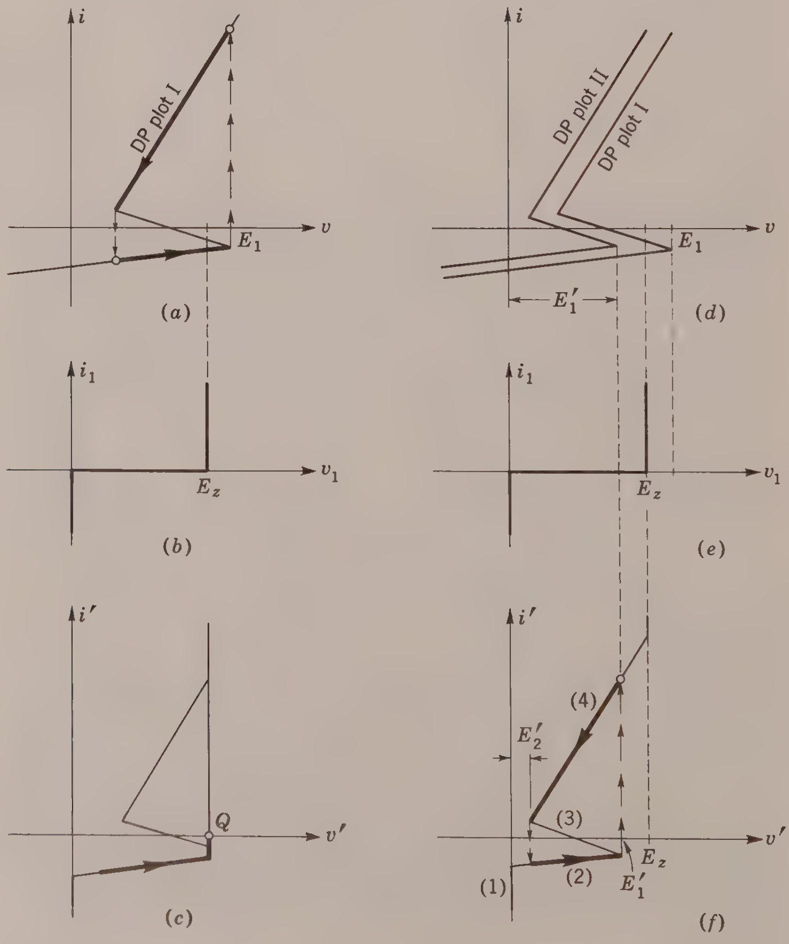


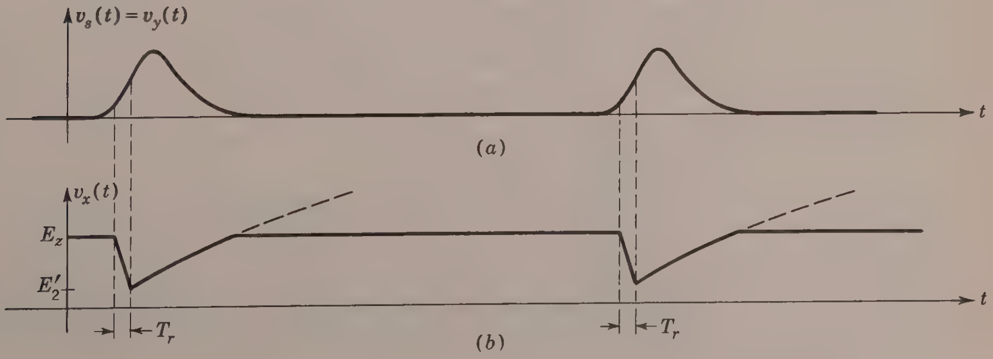
Fig. 17-21. The composite DP plots in the presence and absence of the switching signal show the basic mechanism for initiating a linear sweep.

tained as shown in Fig. 17-21*c*. Observe that point Q is a stable equilibrium state, and hence no switching will occur. In order to initiate the sweep, we must connect a switching voltage source $v_s(t)$ as shown in Fig. 17-20*b*. This has the effect of shifting the original DP Plot I to the new DP Plot II shown in Fig. 17-21*d*. If this shift is large enough so that the new threshold voltage E_1' is smaller than the zener voltage E_z, namely,

$$E_1' < E_z \tag{17-9}$$

then the composite DP plot will be as shown in Fig. 17-21*f*. In this case switching will occur, and a linear sweep will be initiated after the dynamic route has returned to segment 2 of the composite DP plot. Since the sweep will terminate when $v_s(t) = 0$, the input pulse train can be used as the switching signal itself. In this case we can guarantee that the duration of the linear sweep will be approximately equal to the pulse width. For example, if the pulse train $v_s(t)$ is shown in Fig. 17-22*a*, then the output waveform $v_x(t)$ will be as shown in Fig. 17-22*b*. Observe that the linear sweep will occur only after the retrace time T_r is completed. This means that the initial portion of the pulse will not be displayed on the oscilloscope screen. Hence this method of triggering is suitable only for tracing slowly varying waveforms so that only a small initial portion of the pulse is lost. The advantages of this method over the electronic switching method are the inherent simplicity of the circuit and the fact that the same input pulse train can be used as the switching signal, thus obviating the need for additional external circuitry. Another desirable feature of this method is that by connecting a mechanical switch in series with the zener diode, we can obtain either a triggered time-base

Fig. 17-22. The output waveform of a time-base generator and its associated input-triggering pulse train.



generator or a free-running time-base generator by merely closing or opening the switch.

An examination of the circuit shown in Fig. 17-20*b* and the switching mechanism in Fig. 17-21 shows that the effect of the zener diode is to prevent the capacitor voltage from exceeding the zener breakdown voltage E_z. Observe that as soon as the diode "catches" this breakdown voltage, it behaves like a battery of voltage E_z. Because of this the diode (when used for this purpose) is often referred to as a *catching diode.* When a zener diode is not available, the catching diode can be realized by a junction diode and a battery. In view of this "catching" phenomenon, the switching must be initiated before the capacitor voltage reaches the critical value E_z. This is accomplished by applying an appropriate switching signal so that the resulting DP plot will not contain a vertical segment intersecting the v' axis as shown in Fig. 17-21*f*.

Exercise 1: Describe the design of an ideal current time-base generator analogous to that shown in Fig. 17-16.

Exercise 2: Design a triggered time-base generator using a *p*-channel FET switch and a constant-current diode as the nonlinear elements.

17-5 SUMMARY

Time-base generator A first-order dynamic network for generating a linear output voltage waveform (voltage time-base generator) or a linear output current waveform (current time-base generator).

Two types of time-base generators

1. Free-running time-base generator. The output voltage or current is a periodic sawtooth waveform.
2. Triggered time-base generator. The output voltage or current is an aperiodic linear ramp initiated by a triggering signal.

Some practical considerations

1. The retrace time in a practical free-running time-base generator is never zero. It must be made as small as possible.
2. The waveform during each sweep-time interval is not exactly linear. Linearity can be improved by altering the slope of the appropriate DP plot segments.

3. In order to obtain a stationary display in the oscilloscope screen, an external synchronization signal may be applied through an appropriate synchronization circuit. The synchronization signal is usually derived from the time function being displayed. It consists of a periodic string of pulses whose frequency must be an *integral submultiple* of the input signal frequency.

PROBLEMS

17-1 Consider the neon-bulb time-base generator circuit shown in Fig. P17-1*a*, where the v_1-i_1 curve of the neon bulb is shown in Fig. P17-1*b*.

(a) Plot the voltage waveform $v(t)$ at steady state.

(b) Find the sweep time T_s and the retrace time T_r.

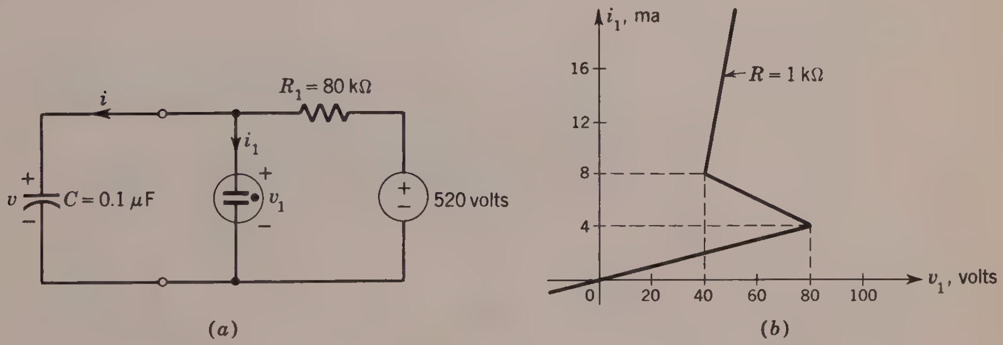


Fig. P17-1.

17-2 A neon-bulb sawtooth generator circuit is shown in Fig. P17-2. Using the v_1-i_1 curve shown in Fig. P17-1*b*, find an appropriate set of values for R, E, and C in order to generate a 1-kHz negative sawtooth voltage waveform.

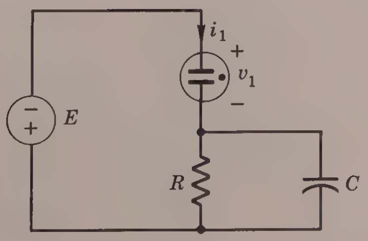
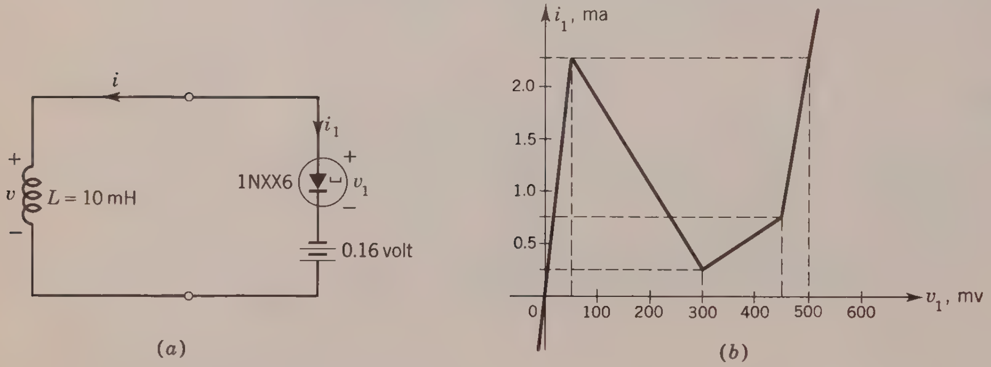


Fig. P17-2.

17-3 A relatively simple current time-base generator circuit is shown in Fig. P17-3*a*. The v_1-i_1 curve of the tunnel diode type 1NXX6 is approximated by the piecewise-linear curve shown in Fig. P17-3*b*.

(a) Plot the current waveform $i(t)$ at steady state.

(b) Find the sweep time T_s and the retrace time T_r.



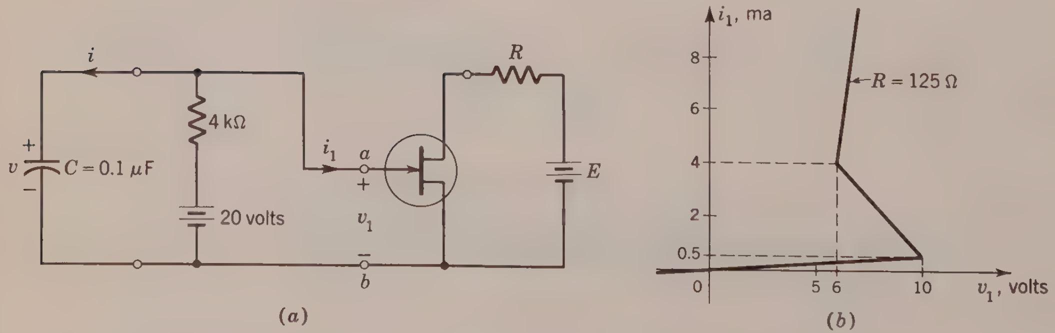
(*a*) (*b*)

Fig. P17-3.

17-4 The unijunction transistor in the circuit shown in Fig. P17-4*a* is biased by R and E so that it exhibits a v_1-i_1 curve across terminals *a*-*b* as shown in Fig. P17-4*b*.

(a) Plot the voltage waveform $v(t)$ at steady state.

(b) Find the sweep time T_s and the retrace time T_r.



(*a*) (*b*)

Fig. P17-4.

17-5 A common transistor time-base generator is the "bootstrap" circuit shown in Fig. P17-5*a*.

(a) Using *n*-*p*-*n* transistor model C_E and assuming $I_{CEO} = 0$ and $V_{CEO} = 0$, show that the DP plot across terminals *a*-*b* is given by Fig. P17-5*b*.

(b) Can segment 2 of the DP plot be made to appear as our ideal current source for a given value of β? If β is variable, can it be made into an ideal current source?

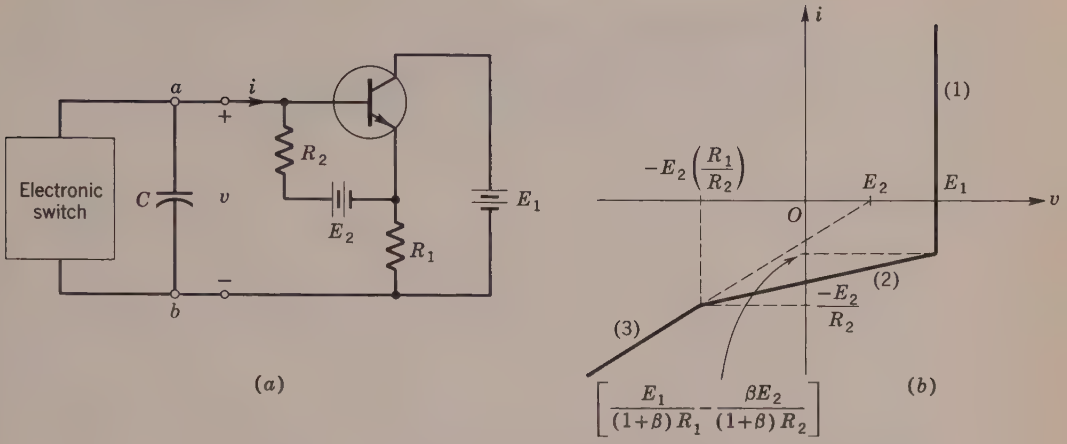

a
i
+
v
−
b
Electronic switch
C
R_2
E_2
R_1
E_1
(a)
i
(1)
$-E_2\left(\frac{R_1}{R_2}\right)$
E_2
E_1
v
O
(2)
$\frac{-E_2}{R_2}$
(3)
$\left[\frac{E_1}{(1+\beta)R_1}-\frac{\beta E_2}{(1+\beta)R_2}\right]$
(b)

Fig. P17-5.

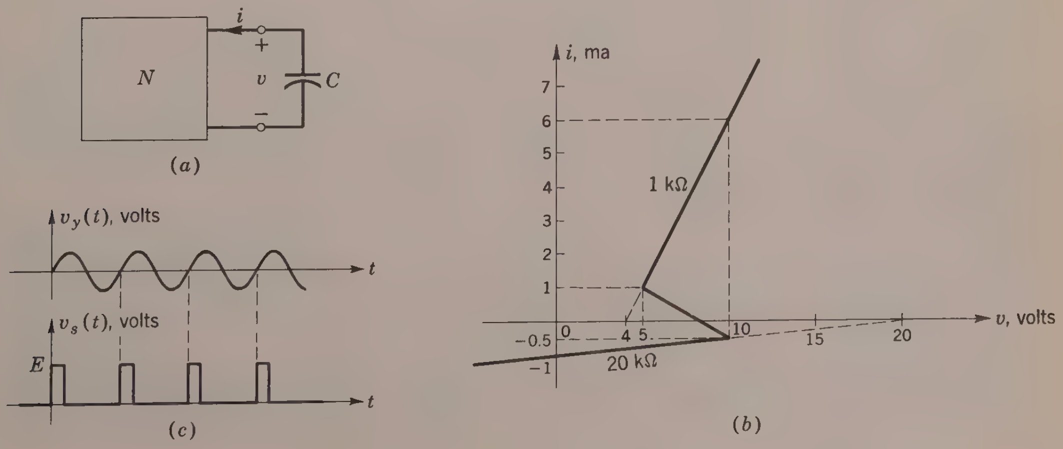
17-6 The circuit shown in Fig. P17-6*a* is to be used as a time-base generator in a cathode-ray oscilloscope. The DP plot of the nonlinear resistive box is given in Fig. P17-6*b*.

(a) Find C such that the frequency of oscillation is 1 kHz and sketch the waveform $v(t)$.

(b) It is desired to synchronize this time-base generator with a 1.1-kHz sine wave. To accomplish this a pulse train $v_s(t)$ is derived from the sine wave as shown in Fig. P17-6*c*. Indicate how $v_s(t)$ should be applied in order to synchronize $v(t)$ with $v_y(t)$. Assuming that E is of sufficient magnitude to accomplish switching instantaneously, sketch a typical dynamic route.

(c) What is the minimum magnitude for E that ensures synchronization?

Fig. P17-6.



18 ANALYSIS OF NONAUTONOMOUS FIRST-ORDER NONLINEAR NETWORKS

18-1 MOTIVATION AND STRATEGIES

All the methods developed in the preceding chapters are based on the fundamental property that over each appropriately chosen time interval the resistive subnetwork can be characterized by a fixed DP plot. This property ceases to be true whenever the network contains ac independent sources described by time functions which are not of a stepwise nature. Based on our earlier definition (Secs. 4-6-2 and 13-4-1), this class of dynamic networks is said to be *nonautonomous.*[1] Roughly speaking, most practical autonomous networks can be considered as *waveform generators* or *oscillators* because they do not require an external driving source, other than stepwise triggering signals. In contrast with this, most practical nonautonomous networks can be considered as *waveshaping networks* or *signal processors* because the time functions generated by the ac sources are being transformed or processed upon passing through such networks.

Nonautonomous networks are commonly encountered in practice for two reasons. First, there are many practical signal processors which are inherently nonautonomous. Second, in order to obtain the transient response of quasi-resistive networks,[2] it is necessary to include major parasitic capacitors and inductors in the analysis.

Since the DP plot (seen by the energy-storage elements) changes from time to time in a nonautonomous network, new approaches are needed to analyze this class of networks. Although the numerical techniques described in Sec. 4-7 are valid for both autonomous and nonautonomous networks, they suffer from the same objections raised earlier in Sec. 14-1; namely, the solutions do not give a qualitative picture of a network's behavior. To overcome this objection, we shall introduce a powerful and general graphical approach, the *constant-slope network approach.* Before we do this, however, we shall consider first two special classes of nonautonomous networks which are amenable to an analytic solution, namely, *linear* and *piecewise-linear networks.*

---

[1] Although the class of nonautonomous networks may include other time-varying elements such as *time-varying resistors, inductors,* and *capacitors,* only time-varying independent sources are allowed in this chapter. Consequently, most of the theorems and methods are formulated for this class of networks and may not be valid for the more general case.

[2] A quasi-resistive network was defined earlier in Sec. 9-1 as a functional network which behaves as a resistive network *most of the time* except during short periods where the waveforms are undergoing rapid changes. A waveform may undergo rapid changes during the application of stepwise signals.

18-2 NONAUTONOMOUS FIRST-ORDER LINEAR NETWORK ANALYSIS

Any nonautonomous first-order linear network can be represented by the black-box configuration shown in Fig. 18-1*a*.[1] The ac-resistive subnetwork N may contain linear resistors, ac or dc independent sources, controlled sources, and other linear resistive elements. Since all resistors are linear and time-invariant, the resistive subnetwork N can be replaced by its Thévenin equivalent circuit as shown in Fig. 18-1*b* or by its Norton equivalent circuit as shown in Fig. 18-1*c*.[2] The resistance R is the equivalent resistance seen by the capacitor or inductor when all *independent* sources inside N are set equal to zero. The time functions $v_i(t)$ and $i_i(t)$ represent, respectively, the equivalent open-circuit voltage and the equivalent short-circuit current across terminals *a-b*. Our object is to be able to calculate $v_o(t)$ corresponding to any $v_i(t)$, and to calculate $i_o(t)$ corresponding to any $i_i(t)$. Since the two circuits are duals, it suffices to consider only the capacitor case.

18-2-1 ZERO-STATE RESPONSE

Without loss of generality let us assume that $v_i(t) = 0$ for all $t \leq t_0$, where t_0 is called the *initial time* and is usually taken to be zero. The pertinent state variable in this case is the capacitor voltage v_o. Recall that the capacitor may have an initial voltage $v_o(t_0)$ even though $v_i(t) = 0$ for $t \leq t_0$. The solution $v_o(t)$ must depend, therefore, on both $v_i(t)$ and the initial state $v_o(t_0)$. As will be shown in the next section, for a linear network the effect of $v_i(t)$ and $v_o(t_0)$ on the solution $v_o(t)$ can be treated separately. Hence, in this section we shall find $v_o(t)$ corresponding to the *zero initial state* $v_o(t_0) = 0$. To emphasize that this is not the complete solution, we shall call $v_o(t)$ corresponding to $v_o(t_0) = 0$ the *zero-state response.*

The normal-form equation for the network in Fig. 18-1*b* is given by

$$\frac{dv_o}{dt} = -\frac{1}{\tau}v_o + \frac{1}{\tau}v_i(t) \qquad (18\text{-}1)$$

---

[1] This assumes that the network contains neither loops of capacitors and voltage sources nor cut sets of inductors and current sources.

[2] We have arbitrarily chosen the energy-storage element in Fig. 18-1*b* to be a capacitor and in Fig. 18-1*c* to be an inductor. Obviously, the equivalent circuit is applicable to both cases.

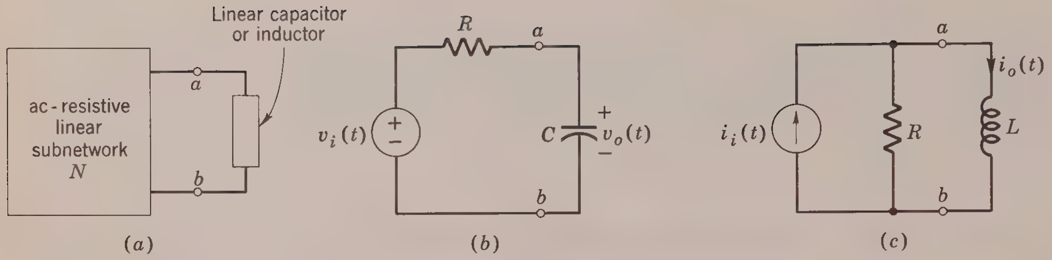

Linear capacitor or inductor
ac-resistive linear subnetwork N
a
b
R
a
$v_i(t)$
+
−
C
+
$v_o(t)$
−
b
a
$i_o(t)$
$i_i(t)$
R
L
b
(a)
(b)
(c)

Fig. 18-1. Any nonautonomous first-order *linear* network can be represented by an equivalent Thévenin or Norton network.

where $\tau \equiv RC$ is the time constant of the circuit. We shall now derive the solution in the form of a theorem.

ZERO-STATE RESPONSE THEOREM

The zero-state response of any nonautonomous first-order capacitor-loaded linear network is given by[1]

$$v_o(t) = \frac{1}{\tau} e^{-t/\tau} \int_{t_0}^{t} e^{y/\tau} v_i(y)\, dy \qquad t \geq t_0 \tag{18-2}$$

To prove this theorem, it suffices to show that Eq. (18-2) has a zero initial state and that it satisfies Eq. (18-1). First, we observe that

$$v_o(t_0) = \frac{1}{\tau} e^{-t_0/\tau} \int_{t_0}^{t_0} e^{y/\tau} v_i(y)\, dy = 0$$

since the upper and the lower limits of integration are equal. Hence, the initial state is indeed zero. Next, let us differentiate Eq. (18-2) with respect to the variable t:

$$\begin{aligned} \frac{dv_o(t)}{dt} &= \left[-\frac{1}{\tau^2} e^{-t/\tau}\right]\left[\int_{t_0}^{t} e^{y/\tau} v_i(y)\, dy\right] + \left[\frac{1}{\tau} e^{-t/\tau}\right][e^{t/\tau} v_i(t)] \\ &= \frac{1}{\tau} v_i(t) - \frac{1}{\tau^2} e^{-t/\tau} \int_{t_0}^{t} e^{y/\tau} v_i(y)\, dy \end{aligned} \tag{18-3}$$

Substituting Eqs. (18-2) and (18-3) into Eq. (18-1) reduces both sides of this equation to an identity. This proves that Eq. (18-2) is indeed the zero-state response.[2]

---

[1] The variable y in Eq. 18-2 is a dummy variable and may be replaced by any other symbol.

[2] We can rewrite Eq. (18-2) into the form

$$v_o(t) = \int_{t_0}^{t} \left[\frac{1}{\tau} e^{-(t-y)/\tau}\right] v_i(y)\, dy$$

or

$$v_o(t) = \int_{t_0}^{t} h(t-y) v_i(y)\, dy$$

where $h(t) \equiv (1/\tau)e^{-t/\tau}$. The second integral is usually called the *convolution* (Fredholm, superposition) *integral* in mathematics. It can be shown that the zero-state response of *any* nonautonomous nth-order linear network is always a convolution integral with an appropriate $h(t)$, called the *impulse response*.

EXAMPLES

1. Let $v_i(t)$ be the *step function* shown earlier in Fig. 15-5*a*. In this case $v_i(t) = 0$ for $t \leq 0$ and $v_i(t) = E$ for $t > 0$. Hence, we can substitute $t_0 = 0$ and $v_i(y) = E$ in Eq. (18-2) to obtain

$$v_o(t) = \frac{1}{\tau} e^{-t/\tau} \int_0^t e^{y/\tau}(E)\, dy = E[1 - e^{-t/\tau}] \qquad t \geq 0 \qquad (18\text{-}4)$$

   This solution, of course, agrees with that shown in Fig. 15-6, as it should.

2. Let $v_i(t)$ be a *unit-ramp* function; that is,

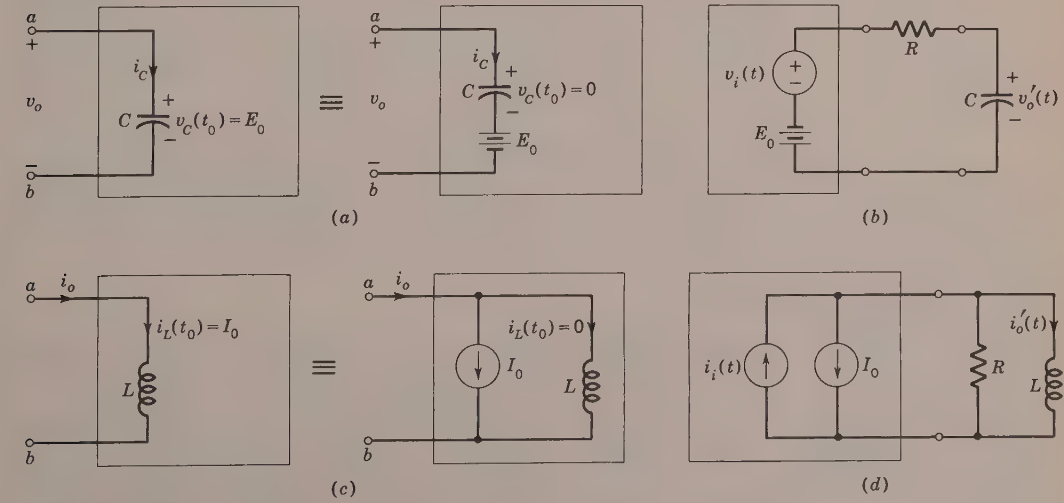
$$v_i(t) = \begin{cases} 0 & t \leq 0 \\ t & t > 0 \end{cases} \qquad (18\text{-}5)$$

   In this case, $t_0 = 0$, and $v_i(y) = y$. Hence, Eq. (18-2) becomes

$$v_o(t) = \frac{1}{\tau} e^{-t/\tau} \int_0^t e^{y/\tau} y\, dy = t - \tau(1 - e^{-t/\tau}) \qquad t \geq 0 \qquad (18\text{-}6)$$

18-2-2 COMPLETE RESPONSE

Let us now consider how to handle the case where the initial state is not zero; i.e., suppose $v_o(t_0) = E_0$ and $i_o(t_0) = I_0$. Consider first the two black boxes shown in Fig. 18-2*a*. The two capacitors are assumed to have identical capacitance C. However, the one on the left has an initial voltage $v_C(t_0) = E_0$, while the one on the right



(*a*) (*b*)

(*c*) (*d*)

Fig. 18-2. Any linear capacitor (inductor) with an initial state $E_0(I_0)$ can be replaced by a linear capacitor (inductor) with zero initial state in series (parallel) with a voltage (current) source.

has zero initial state. These two black boxes are equivalent because the voltages $v_o(t)$ across terminals *a-b* are both given by

$$v_o(t) = E_0 + \frac{1}{C}\int_{t_0}^{t} i_C(t)\,dt \tag{18-7}$$

Hence, if the capacitor in Fig. 18-1*b* has an initial state $v_o(t_0) = E_0$, we can replace it by a zero-initial-state capacitor as shown in Fig. 18-2*b*. The voltage sources $v_i(t)$ and E_0 can be combined into an equivalent voltage source $v_i'(t) = v_i(t) - E_0$. The zero-state response given by Eq. (18-2) can be used to calculate $v_o'(t)$. The *complete response* is, therefore, obtained upon adding $v_o'(t)$ with E_0; namely,

$$v_o(t) = E_0 + \frac{1}{\tau}e^{-t/\tau}\int_{t_0}^{t} e^{y/\tau}[v_i(y) - E_0]\,dy \qquad t \geq t_0 \tag{18-8}$$

This equation can be simplified into the form

$$v_o(t) = \underbrace{[E_0e^{-(t-t_0)/\tau}]}_{\substack{\text{Zero-input}\\ \text{response}}} + \underbrace{\left[\frac{1}{\tau}e^{-t/\tau}\int_{t_0}^{t} e^{y/\tau}v_i(y)\,dy\right]}_{\text{Zero-state response}} \qquad t \geq t_0 \tag{18-9}$$

An examination of Eq. (18-9) shows that the complete response consists of the sum of two parts. The first part can be interpreted as the response when the input $v_i(t) = 0$, and hence it is called the *zero-input response*.[1] The second part is identical with Eq. (18-2) and is, therefore, the zero-state response.

By a dual development it is clear that a linear inductor with an initial current I_0 is equivalent to one with zero initial current connected in parallel with a current source as shown in Fig. 18-2*c*. Hence, if the inductor in Fig. 18-1*c* has a nonzero initial state, it can be reduced to the equivalent network shown in Fig. 18-2*d*. The complete response is, therefore, given by the following dual of Eq. (18-9):

$$i_o(t) = \underbrace{[I_0e^{-(t-t_0)/\tau}]}_{\substack{\text{Zero-input}\\ \text{response}}} + \underbrace{\left[\frac{1}{\tau}e^{-t/\tau}\int_{t_0}^{t} e^{y/\tau}i_i(y)\,dy\right]}_{\text{Zero-state response}} \qquad t \geq t_0 \tag{18-10}$$

We shall now summarize the above results in a theorem.

---

[1] Notice that if the time constant τ is positive, the zero-input response decays to zero in about five time constants. Therefore, after this initial transient time interval, the complete response reduces to the zero-state response, and the network is usually said to be in *steady state*. In contrast to this, the initial transient portion of the complete response is usually called the *transient response*.

COMPLETE-RESPONSE THEOREM

The solution of any nonautonomous first-order linear network corresponding to a given initial state is always given by the sum of the zero-input response and the zero-state response (Eq. 18-9 for the capacitor-loaded case and Eq. 18-10 for the inductor-loaded case). The zero-input response is the solution due only to the initial condition (i.e., with all independent sources set equal to zero). The zero-state response is the solution with zero initial condition.

18-2-3 VERY LARGE AND VERY SMALL TIME-CONSTANT NETWORKS

There are many practical circuits wherein the time constant τ is either very large or very small. In such cases it is sometimes more instructive to obtain the solution in an alternate form.

Consider first the capacitor network shown in Fig. 18-1*b*. We shall now show that the zero-state response of this network is given by

$$\begin{aligned} v_o(t) = {} & \frac{1}{\tau}\int_{t_0}^{t} v_i(y)\,dy - \frac{1}{\tau^2}\int_{t_0}^{t}\int_{t_0}^{y_2} v_i(y_1)\,dy_1\,dy_2 \\ & + \frac{1}{\tau^3}\int_{t_0}^{t}\int_{t_0}^{y_3}\int_{t_0}^{y_2} v_i(y_1)\,dy_1\,dy_2\,dy_3 + \cdots + (-1)^{k+1}\frac{1}{\tau^k}\int_{t_0}^{t}\int_{t_0}^{y_k}\cdots \\ & \int_{t_0}^{y_2} v_i(y_1)\,dy_1\,dy_2\cdots dy_k + \cdots \qquad t \geq t_0 \qquad (18\text{-}11) \end{aligned}$$

To prove that Eq. (18-11) is indeed the zero-state response, we first observe that $v_o(t_0) = 0$, and hence the initial state is zero. Next, let us differentiate both sides of Eq. (18-11) with respect to the variable t.

$$\begin{aligned} \frac{dv_o(t)}{dt} = {} & \frac{1}{\tau}v_i(t) - \frac{1}{\tau^2}\int_{t_0}^{t} v_i(y_1)\,dy_1 + \frac{1}{\tau^3}\int_{t_0}^{t}\int_{t_0}^{y_2} v_i(y_1)\,dy_1\,dy_2 \\ & + \cdots + (-1)^{k+1}\frac{1}{\tau^k}\int_{t_0}^{t}\int_{t_0}^{y_{k-1}}\cdots\int_{t_0}^{y_2} v_i(y_1)\,dy_1\,dy_2\cdots dy_{k-1} + \cdots \end{aligned} \qquad (18\text{-}12)$$

If we substitute Eqs. (18-11) and (18-12) into Eq. (18-1), we find that all terms in the left-hand side cancel out with the terms on the right-hand side. This establishes the proof.

Since the zero-state response given by Eq. (18-11) contains an infinite number of terms, it is difficult to show that the series gives exactly the same solution as Eq. (18-2). At first sight, Eq. (18-11) might appear to be a circuitous if not impractical form of solution. This is true except when the time constant τ is very large. Observe that since the coefficient in the kth term of the series is given by $1/\tau^k$, it is clear that if $\tau \gg 1$, then $1/\tau^k \approx 0$. Hence, if we rewrite Eq. (18-11) into the form

$$v_o(t) = \frac{1}{\tau}\int_{t_0}^{t} v_i(y)\,dy + \epsilon(t) \tag{18-13}$$

where

$$\epsilon(t) \equiv \sum_{k=2}^{\infty} (-1)^{k+1}\frac{1}{\tau^k}\int_{t_0}^{t}\int_{t_0}^{y_k}\cdots\int_{t_0}^{y_2} v_i(y_1)\,dy_1\,dy_2\cdots dy_k \tag{18-14}$$

then if $\tau \gg 1$, we have

$$v_o(t) \approx \frac{1}{\tau}\int_{t_0}^{t} v_i(y)\,dy \tag{18-15}$$

and

$$\epsilon(t) \approx 0 \tag{18-16}$$

Equation (18-15) shows that the output voltage is proportional to the time integral of the input voltage. The circuit shown in Fig. 18-1*b* can, therefore, be used as a voltage integrator, provided the time constant is made very large.[1] An inspection of Eq. (18-15) shows that the output voltage will become very small if τ is made very large. Hence, in practice a compromise must be made, and the first few terms in Eq. (18-14) will give a very accurate measure of the error or departure from an ideal integrator. Observe that if $v_i(t)$ does not change sign as t increases, then the error $\epsilon(t)$ given by Eq. (18-14) also increases with time. This means that the simple integrator circuit shown in Fig. 18-1*b* will perform satisfactorily only during a finite interval of time.

Consider next the circuit shown in Fig. 18-3*a*. It can be proved by the same method of substitution as before that the zero-state response in this case is given by

$$v_o(t) = \tau\frac{dv_i}{dt} - \tau^2\frac{d^2v_i}{dt^2} + \tau^3\frac{d^3v_i}{dt^3} - \cdots + (-1)^{k+1}\tau^k\frac{d^kv_i}{dt^k} + \cdots \tag{18-17}$$

If we rewrite this in the form

$$v_o(t) = \tau\frac{dv_i}{dt} + \epsilon(t) \tag{18-18}$$

---

[1] A more expensive type of integrator using operational amplifiers was given earlier in Fig. 3-23.

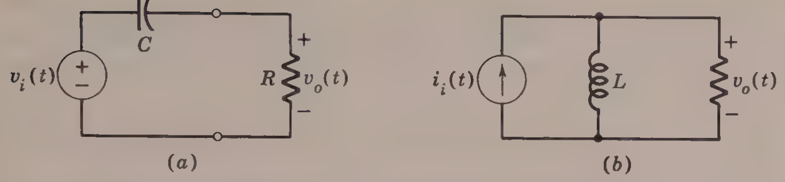


Fig. 18-3. Two simple differentiator circuits with very small time constants.

where

$$\epsilon(t) = -\tau^2 \frac{d^2 v_i}{dt^2} + \tau^3 \frac{d^3 v_i}{dt^3} - \cdots + (-1)^{k+1} \tau^k \frac{d^k v_i}{dt^k} + \cdots \tag{18-19}$$

then it is clear that as $\tau \to 0$, $\epsilon(t) \to 0$. Hence, for a very small time constant τ we can write

$$v_o(t) \approx \tau \frac{dv_i}{dt} \tag{18-20}$$

and

$$\epsilon(t) \approx 0 \tag{18-21}$$

Equation (18-20) shows that the circuit in Fig. 18-3*a* can be used as a voltage differentiator, provided the time constant is made very small.[1] Similarly, the dual circuit shown in Fig. 18-3*b* can be used as a current differentiator.

Exercise 1: Prove that the zero-state response of any nonautonomous first-order inductor-loaded linear network is given by

$$i_o(t) = \frac{1}{\tau} e^{-(t/\tau)} \int_{t_0}^{t} e^{y/\tau} i_i(y)\, dy$$

where $\tau \equiv L/R$.

Exercise 2: Find the zero-state response $v_o(t)$ of the network in Fig. 18-1*b* corresponding to

$$v_i(t) = \begin{cases} 0 & t \leq -2 \\ k(t+2) & t > -2 \end{cases}$$

Exercise 3: Repeat Exercise 2 for

$$v_i(t) = \begin{cases} 0 & t \leq 0 \\ kt^2 & t > 0 \end{cases}$$

Exercise 4: Find the complete response of a capacitor-loaded network with $R = C = 1$; $v_i(t) = 0$, $t \leq 0$, $v_i(t) = t$, $t > 0$; and $v_o(0) = 10$.

---

[1] A more expensive type of differentiator using operational amplifiers was given earlier in Fig. 3-22.

Exercise 5: Find the complete response of an inductor-loaded network with $R = 1\ \text{k}\Omega$, $L = 2$ mH; $i_i(t) = 0$, $t \leq 0$, $i_i(t) = t^2$ ma, $t > 0$, and $i_i(0) = 5$ ma.

Exercise 6: Explain why the name *transient response* is misleading when the time constant is negative.

Exercise 7: Show how the circuit in Fig. 18-1*c* can be designed to function as a current integrator.

Exercise 8: Verify that Eq. (18-17) is indeed the zero-state response of the network shown in Fig. 18-3*a*.

Exercise 9: Discuss the effect of the time constant on the practical design of a differentiator using the circuits shown in Fig. 18-3.

Exercise 10: Replace the capacitor in Fig. 18-1*b* by an inductor and show that the zero-state response $v_o(t)$ across the inductor is given by Eq. (18-17). Discuss how a voltage differentiator can be designed using this circuit.

Exercise 11: Replace the capacitor in Fig. 18-3*a* by an inductor and show that the zero-state response $v_o(t)$ across the resistor is given by Eq. (18-11). Discuss how a voltage integrator can be designed using this circuit.

Exercise 12: Exhibit some illustrative examples showing that the effect of the integrator is to *smooth out* the input waveform, whereas the effect of the differentiator is to *sharpen* the input waveform.

18-3 NONAUTONOMOUS FIRST-ORDER PIECEWISE-LINEAR NETWORK ANALYSIS

The complete response derived in the preceding section (Eqs. 18-9 and 18-10) can be combined with the iterative piecewise-linear method developed in Chap. 12 to analyze any nonautonomous first-order network, provided the resistive subnetwork is in the canonic form (see Sec. 12-1). The method is based on the observation that if the ac independent sources are represented by continuous time functions, then during each appropriately chosen time interval (t_j,t_{j+1}), the operating point of each nonlinear resistor will remain on the same segment. Hence, during each time interval (t_j,t_{j+1}), the network is linear and may be reduced into the equivalent form shown in Fig. 18-1*b* or *c*. It must be emphasized, however, that the values of R and $v_i(t)$ in Fig. 18-1*b* and the values of R and $i_i(t)$ in Fig. 18-1*c* are generally different over different time intervals. Once the pertinent values of R, $v_i(t)$, and $i_i(t)$ are assigned, the solution corresponding to a given initial condition can be found simply by using Eq. 18-9 for the capacitor-loaded case and Eq. 18-10 for the inductor-loaded case. The fundamental problem that we must solve, therefore, consists of finding out how to subdivide the time axis into appropriate intervals (t_j,t_{j+1}). We shall develop this technique for two typical cases. The extension to the general case will become obvious. In order to avoid unnecessary complexities which do not affect the validity of the method, we shall state the following standing assumption:

UNIQUENESS HYPOTHESIS

All networks considered in this section are assumed to possess unique solutions at all times.

18-3-1 ONE NONLINEAR RESISTOR CASE

The simplest type of a piecewise-linear network consists of a nonlinear resistor in series or in parallel with a linear ac-resistive subnetwork N as shown in Fig. 18-4*a* and *b*, respectively. As usual, N may contain any number of linear elements, ac independent sources, and controlled sources. In any case we can replace the network N by its Thévenin equivalent circuit and the nonlinear resistor by its iterative Thévenin equivalent as shown in Fig. 18-4*c*. The dual case corresponding to Fig. 18-4*b* is shown in Fig. 18-4*d*.

It suffices to consider the capacitor-loaded case. Let us assume that the nonlinear resistor R_1 is characterized by n piecewise-linear segments. Moreover, in view of the uniqueness hypothesis, the v_1-i_1 curve must be voltage-controlled. This means that the voltage intervals of definition $({}_kE_1^-, {}_kE_1^+)$ of each segment k do not

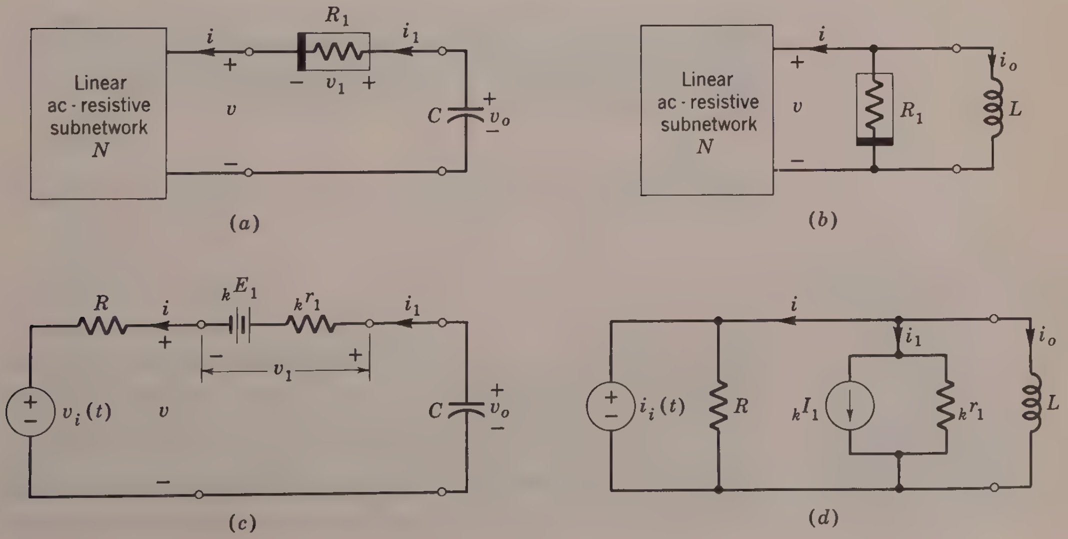


Fig. 18-4. Two equivalent iterative piecewise-linear networks containing one nonlinear resistor.

overlap. Since the variables for deciding which segment is operating at a particular time t_0 are the input voltage $v_i(t_0)$ and the output voltage $v_o(t_0)$ at $t = t_0$, let us derive a relationship for v_1 in terms of v_i and v_o. This is easily found to be

$$v_1 = \left(\frac{-{}_kr_1}{R + {}_kr_1}\right) v_i + \left(\frac{{}_kr_1}{R + {}_kr_1}\right) v_o + \left(\frac{R}{R + {}_kr_1}\right) {}_kE_1 \tag{18-22}$$

Observe that this equation is of the general form[1]

$$v_1 = {}_kH_{11}(v_i) + {}_kH_{12}(v_o) + {}_kK_1 \tag{18-23}$$

where ${}_kH_{11} \equiv -{}_kr_1/(R + {}_kr_1)$, ${}_kH_{12} \equiv {}_kr_1/(R + {}_kr_1)$, and ${}_kK_1 \equiv R/(R + {}_kr_1){}_kE_1$. In order for the network to operate on segment k, which has an interval of definition

$${}_kE_1^- < v_1 < {}_kE_1^+ \tag{18-24}$$

let us substitute Eq. (18-22) in place of v_1 in Eq. (18-24) and simplify to obtain

$$\left(\frac{-{}_kr_1}{R + {}_kr_1}\right) v_i + \left(\frac{{}_kr_1}{R + {}_kr_1}\right) v_o < {}_kE_1^+ - \left(\frac{R}{R + {}_kr_1}\right) {}_kE_1 \tag{18-25}$$

$$\left(\frac{-{}_kr_1}{R + {}_kr_1}\right) v_i + \left(\frac{{}_kr_1}{R + {}_kr_1}\right) v_o > {}_kE_1^- - \left(\frac{R}{R + {}_kr_1}\right) {}_kE_1 \tag{18-26}$$

These two inequalities can be interpreted as the region bounded between two unity-slope straight lines as shown in Fig. 18-5*a*. The voltage intercepts ${}_kE_0^+$ and ${}_kE_0^-$ corresponding to segment k are given, respectively, by

$${}_kE_0^+ = \frac{(R + {}_kr_1){}_kE_1^+ - R({}_kE_1)}{{}_kr_1} \tag{18-27}$$

$${}_kE_0^- = \frac{(R + {}_kr_1){}_kE_1^- - R({}_kE_1)}{{}_kr_1} \tag{18-28}$$

We shall call each of these two boundary lines a *switching line*. Corresponding to each segment k of the v_1-i_1 curve, two switching lines are obtained by simply using Eqs. (18-27) and (18-28).[2] In all cases the switching lines are parallel to each other and have unity slopes. Hence, it is a trivial matter to subdivide the v_i-v_o plane into n parallel strips corresponding to the n segments of the v_1-i_1 curve.

[1] The reader should recognize that this linear form is the two-dimensional extension of Eq. (12-34).

[2] Since ${}_1E_1^- = -\infty$ for the leftmost segment and ${}_nE_1^+ = \infty$ for the rightmost segment, these two switching lines are located at infinity and need not be drawn.

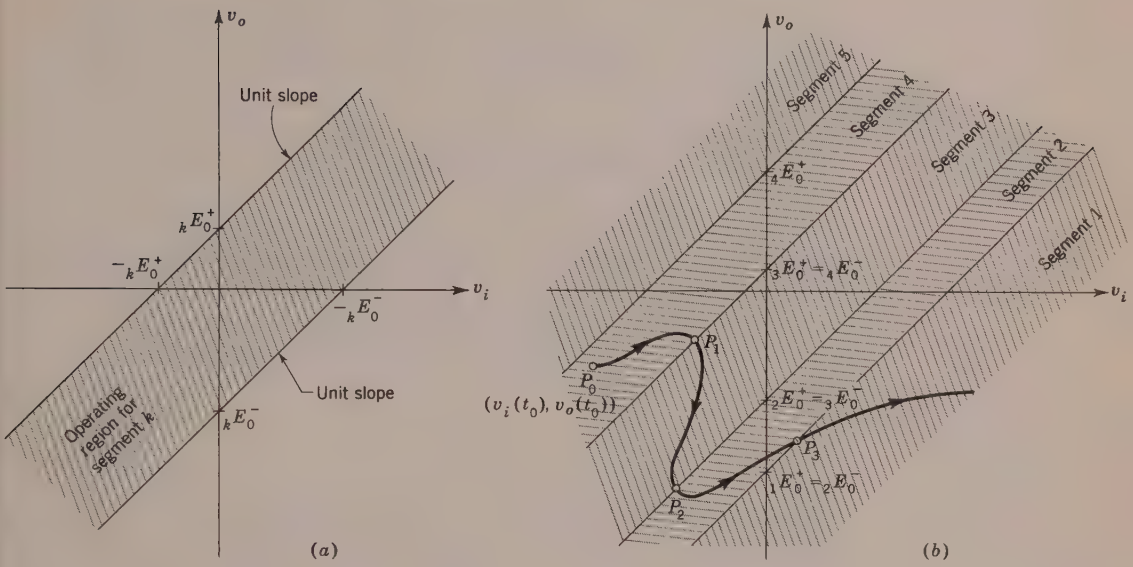


Fig. 18-5. Switching plane defining the linear regions for which the nonlinear resistor is operating at a given segment.

We shall call the subdivided plane the *switching plane.* An example of a switching plane corresponding to a five-segment voltage-controlled v_1-i_1 curve is shown in Fig. 18-5.

Once the switching plane is obtained, it is a simple matter to find the complete response corresponding to any given initial state $v_o(t_0)$ and any input voltage $v_i(t)$. The steps are as follows:

1. Locate the initial point P_0 with coordinates $(v_i(t_0),v_o(t_0))$ on the switching plane. This immediately locates the pertinent initial segment. For example, corresponding to the point P_0 in Fig. 18-5*b*, the network, we know, is operating initially on segment 4.
2. Assign the numerical values for ${}_kE_1$ and ${}_kr_1$ in Fig. 18-4*c* corresponding to the pertinent operating segment obtained from step 1.
3. Reduce the network into the equivalent form shown in Fig. 18-1*b* and obtain the complete response $v_o(t)$ [with initial state $v_o(t_0)$] by using Eq. (18-9).
4. Draw the locus of $(v_i(t),v_o(t))$ on the switching plane until it intersects a switching line at some point P_1.†
5. Identify the time parameter t_1 at P_1 and let P_1 be the new initial point with initial state $v_o(t_1)$.

† This locus can be interpreted as a Lissajous figure and is analogous to the trajectory in the phase plane.

6. Replace the parameter values for ${}_kE_1$ and ${}_kr_1$ by the pertinent segment for $t > t_1$ (segment 3 in the example in Fig. 18-5*b*) and repeat steps 3 to 5.
7. The above procedure is iterated as many times as necessary until the complete response is found for all times $t \geq t_0$.

The above procedure is quite easy to implement in practice. Moreover, observe that the switching plane is independent of the input voltage $v_i(t)$ as well as of the capacitor voltage $v_o(t)$. Hence, once the switching plane for a given network is obtained, we can easily find the complete response corresponding to any input signals $v_i(t)$ and *any* initial state $v_o(t_0)$.

18-3-2 *M* NONLINEAR RESISTOR CASE

Consider next the case where the resistive subnetwork contains m nonlinear resistors but only one ac independent source as shown in Fig. 18-6*a*.[1] The corresponding iterative linear network is shown in Fig. 18-6*b*. Observe that at any time t, the pertinent operating segment of each nonlinear resistor depends only on the input voltage $v_i(t)$ and the output voltage $v_o(t)$. Clearly, the concept of the switching plane is still applicable here, and the same procedure in the preceding section can be used to find the complete response, provided the switching plane can be found.

The procedure for drawing the switching plane is similar to the case considered above, only this time it must be applied to each of the m nonlinear resistors. In other words, we must derive a relationship for each voltage v_j of R_j in terms of v_i and v_o. Since the iterative network in Fig. 18-6*b* is linear, these relationships will be of the same form as Eq. (18-23); thus

$$\begin{aligned} v_1 &= {}_kH_{11}(v_i) + {}_kH_{12}(v_o) + {}_kK_1 \\ v_2 &= {}_kH_{21}(v_i) + {}_kH_{22}(v_o) + {}_kK_2 \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ v_m &= {}_kH_{m1}(v_i) + {}_kH_{m2}(v_o) + {}_kK_m \end{aligned} \tag{18-29}$$

where the subscript k now stands for a particular segment combination, namely, k_1 for R_1, k_2 for R_2, . . . , k_m for R_m. If we denote the interval of definition for each segment k_j of the v_j-i_j curve by

$${}_kE_j^- < v_j < {}_kE_j^+ \tag{18-30}$$

---

[1] In contrast with this the resistive subnetwork shown in Fig. 18-4*a* and *b* may contain *any* number of ac independent sources but only one nonlinear resistor. More than one ac independent source is allowed there because these sources are all contained within a linear resistive subnetwork and hence can be reduced to a single equivalent source.

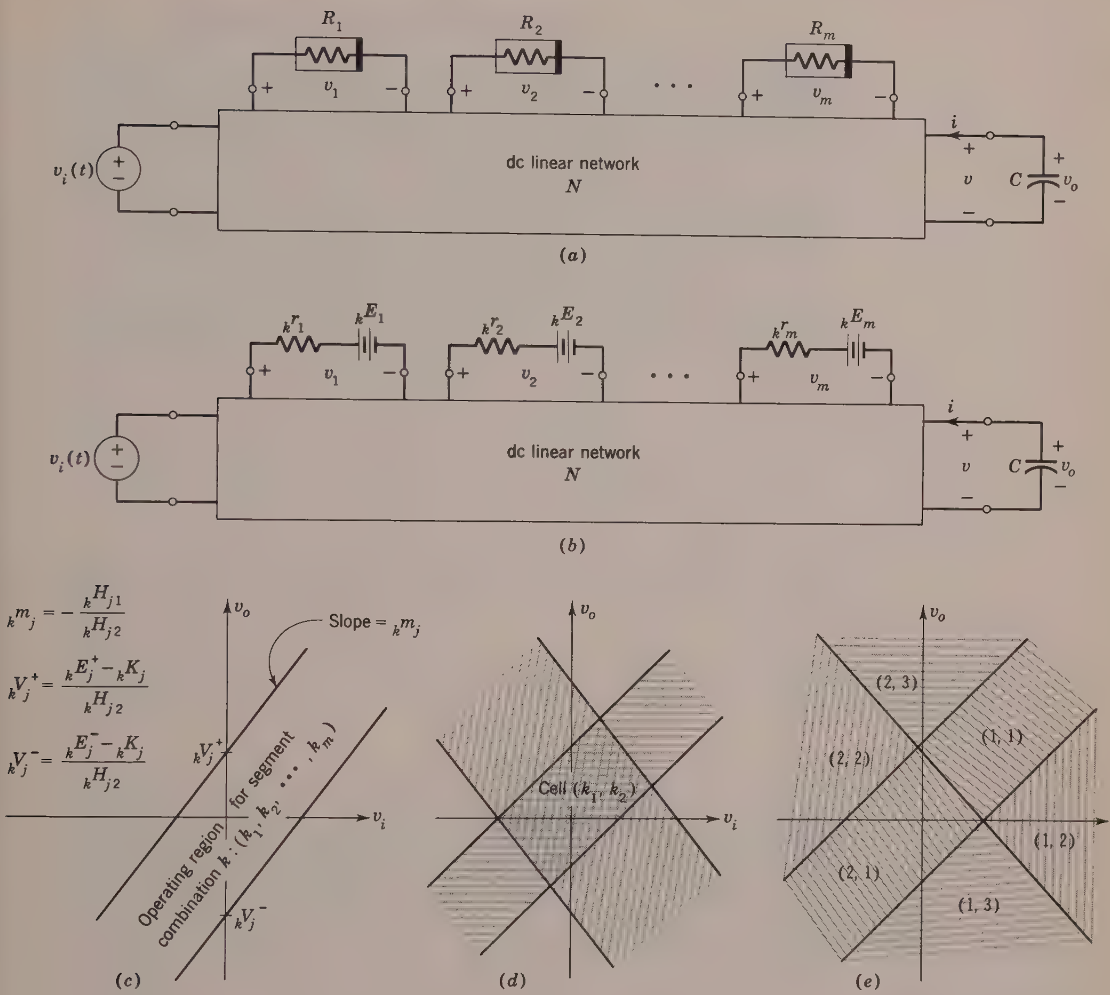


Fig. 18-6. The crucial step for obtaining the complete response of a first-order network with m nonlinear resistors and one ac independent source consists of finding the switching plane of the network.

then each equation in Eq. (18-29) becomes a pair of inequalities similar to Eqs. (18-25) and (18-26); namely,

$$ {}_kH_{j1}(v_i) + {}_kH_{j2}(v_o) < {}_kE_j^+ - {}_kK_j \tag{18-31}$$

$$ {}_kH_{j1}(v_i) + {}_kH_{j2}(v_o) > {}_kE_j^- - {}_kK_j \tag{18-32}$$

where $j = 1, 2, \ldots, m$. As usual, each pair of inequalities can be interpreted geometrically as the area between two parallel straight lines as shown in Fig. 18-6c. Since there are m such strips corresponding to m nonlinear resistors that must be satisfied simultaneously, the area common to all m strips, henceforth called a *cell*, defines the operating region corresponding to the particular segment combination k $(k_1,k_2, \ldots ,k_m)$. An example of a typical cell representing two nonlinear resistors (segment k_1 of R_1 and segment k_2 of R_2) is shown in Fig. 18-6*d*.[1] Just as in the determination of DP plots and TC plots in Secs. 12-3 and 12-4, it usually occurs that some segment combinations do not lead to a common area; in such cases, the corresponding cell does not exist. In general, it is also possible for two or more cells to overlap. However, in view of the uniqueness hypothesis, no overlapping may occur, and one can be sure that the switching plane is filled up with nonoverlapping cells. A typical example of a switching plane with six cells is shown in Fig. 18-6*e*.

The above principles can be generalized to include any number n of ac independent sources. The "cell" k $(k_1,k_2, \ldots ,k_m)$ corresponding to Eqs. (18-31) and (18-32) is now defined by

$$[{}_kH_{j1}(v_1) + {}_kH_{j2}(v_2) + \cdots + {}_kH_{jn}(v_n)] + {}_kH_{j(n+1)}(v_o) < {}_kE_j^+ - {}_kK_j \tag{18-33}$$

$$[{}_kH_{j1}(v_1) + {}_kH_{j2}(v_2) + \cdots + {}_kH_{jn}(v_n)] + {}_kH_{j(n+1)}(v_o) > {}_kE_j^- - {}_kK_j \tag{18-34}$$

where $j = 1, 2, \ldots, m$. Each cell is now represented by a polyhedron in an $(n + 1)$-dimensional space. Hence, instead of a switching plane we now have a *switching space*, and the simplicity of our earlier geometric approach is no longer applicable. It is clear that a computer will be needed to obtain the complete response in this general case.

Exercise 1: Derive the general expression for the complete response $v_o(t)$ of the iterative linear network shown in Fig. 18-4*c* in terms of the parameters R, ${}_kE_1$, ${}_kr_1$, and the initial state $v_o(t_j)$. HINT: Use Eq. (18-9).

Exercise 2: Derive the general expression for the complete response $i_o(t)$ of the iterative linear network shown in Fig. 18-4*d* in terms of the parameters R, ${}_kI_1$, ${}_kr_1$, and the initial state $i_o(t_j)$.

Exercise 3: Derive the equations for drawing the switching plane of the inductor-loaded network shown in Fig. 18-4*d*.

Exercise 4: Derive the equations for drawing the switching plane of an inductor-loaded network containing m nonlinear resistors and one ac independent current source.

---

[1] It is obvious from an inspection of Fig. 18-6*d* that the cell must always assume the form of a polygon, one of whose sides may be at infinity (corresponding to leftmost or rightmost segments).

18-4 THE CONSTANT-SLOPE NETWORK APPROACH

We shall now develop a very general and powerful graphical technique for obtaining the complete response of *any* nonautonomous first-order network. The network may contain any number of nonlinear resistors, ac independent sources, and controlled sources. The energy-storage element is allowed to be nonlinear. Moreover, the characteristic curves need not be represented in piecewise-linear form. The only assumption we impose is that the network has a unique solution for all times and for any initial states. Under this assumption, the normal-form equation is single-valued and is given by

$$\frac{dx}{dt} = f(x,t) \tag{18-35}$$

Any solution $x = X(t)$ of Eq. (18-35) corresponding to an initial state $x(t_0) = x_0$ is a curve in the t-x plane which starts at the point (t_0,x_0) and which satisfies Eq. (18-35) for all times $t \geq t_0$. This means that at any time $t = t_j$, the slope $m(t_j)$ of this solution curve must be given precisely by

$$m(t_j) \equiv \left.\frac{dx}{dt}\right|_{t=t_j} = f(x,t_j) \tag{18-36}$$

as shown in Fig. 18-7*a*. Since Eq. (18-35) defines uniquely the slope at any point on the t-x plane which a solution curve passing through this point must possess, it is clear that if we compute the slope of a sufficient number of points near the solution curve and denote each slope by a short segment, called a *director,* we obtain the system of directors shown in Fig. 18-7*b*. Observe that if we draw a smooth curve starting at P_0 so that it is tangent at each time t_j to the director at the corresponding point, we obtain the

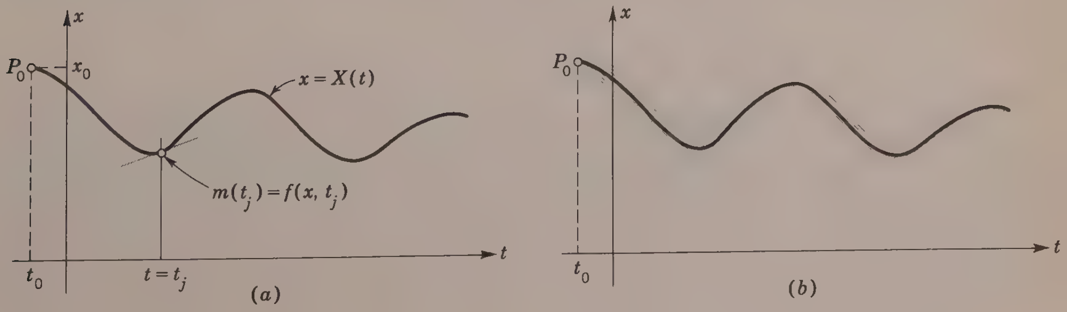


Fig. 18-7. The solution curve (complete response) corresponding to any initial state P_0 can be obtained by passing a smooth curve through P_0 and tangent to the system of directors.

original solution $x = X(t)$ in Fig. 18-7*a*. In view of this observation, it is clear that if we fill up the entire t-x plane with a sufficient number of directors, then the solution waveform corresponding to any initial state can be trivially sketched. In other words, instead of finding only one complete response (solution curve) at a time, we obtain an entire picture of all possible solution curves in one shot. For convenience, we shall refer to the system of directors plotted in accordance with Eq. (18-35) as the *vector field* associated with Eq. (18-35). Since the elements in a vector field are vectors, each director must be interpreted as a *vector* pointing towards the direction of increasing time, namely, *from left to right.*

18-4-1 METHODS FOR DRAWING THE VECTOR FIELD

Conceptually, the simplest method for drawing the vector field is to draw a system of horizontal and vertical lines, thereby forming a grid. The slope at each intersection point is then calculated from Eq. (18-35) and the director drawn. We shall call this the *brute force method.*[1]

A more efficient method for drawing the vector field consists of finding the locus of all points on the t-x plane which have identical slopes. This locus is easily obtained by plotting

$$f(x,t) = m \tag{18-37}$$

where m is the desired slope. Equation (18-37) represents a curve in the t-x plane and is called a *constant-slope curve.* Corresponding to each constant-slope curve $\Gamma(m)$ of slope m, a system of directors with slope m along $\Gamma(m)$ can immediately be drawn. Hence, by assigning a set of convenient values for m, we can plot the corresponding constant-slope curves, and the vector field can be more systematically obtained. Let us illustrate this procedure by an example.

EXAMPLE

Consider the R-C network shown in Fig. 18-1 with $R = 1\ \Omega$ and $C = 1$ F. Let us find the complete response due to the unit-ramp function defined earlier by Eq. (18-5). The normal-form equation is obtained by substituting $\tau = RC = 1$ and $v_i(t) = t$ in Eq. (18-1).

$$\frac{dv_o}{dt} = -v_o + t \qquad t \geq 0 \tag{18-38}$$

---

[1] This brute force method can be easily implemented on a computer in conjunction with a mechanical plotter such as the Calcomp plotter.

The constant-slope curve defined by Eq. (18-37) becomes

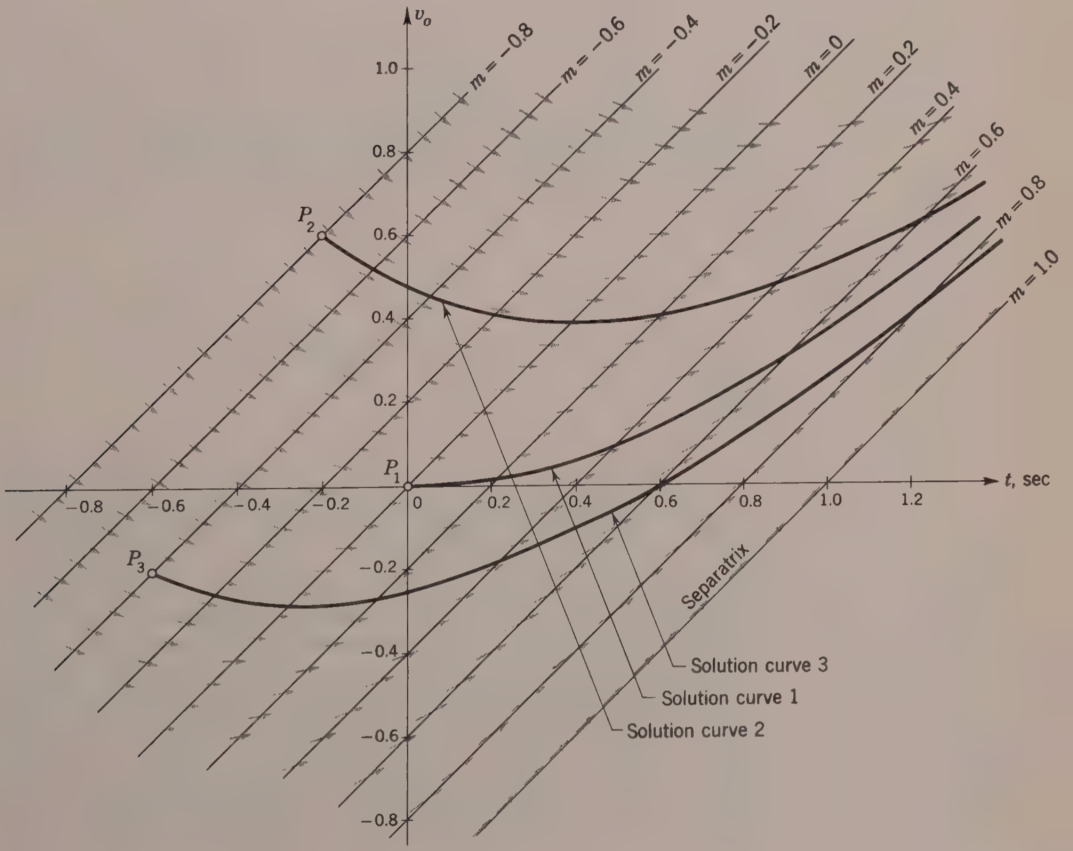
$$f(v_o) = -v_o + t = m$$

or

$$v_o = t - m \tag{18-39}$$

Corresponding to each value m, Eq. (18-39) represents a unit-slope straight line in the t-v_o plane. The vector field corresponding to this network is, therefore, easily drawn as shown in Fig. 18-8. From this vector field we can sketch the complete response corresponding to *any* initial state $v_o(t_0)$. For example, corresponding to the initial state $P_1(0,0)$, the solution curve 1 shown in Fig. 18-8 is obtained. Observe that if Eq. (18-38) is defined also for $t < 0$ (corresponding to a different input time function), the vector field in Fig. 18-8 remains unchanged. Hence, corresponding to the initial states $P_2(-0.2,0.6)$ and $P_3(-0.6,-0.2)$, the solution curves 2 and 3

Fig. 18-8. The vector field as obtained by the constant-slope curves and the associated complete responses.



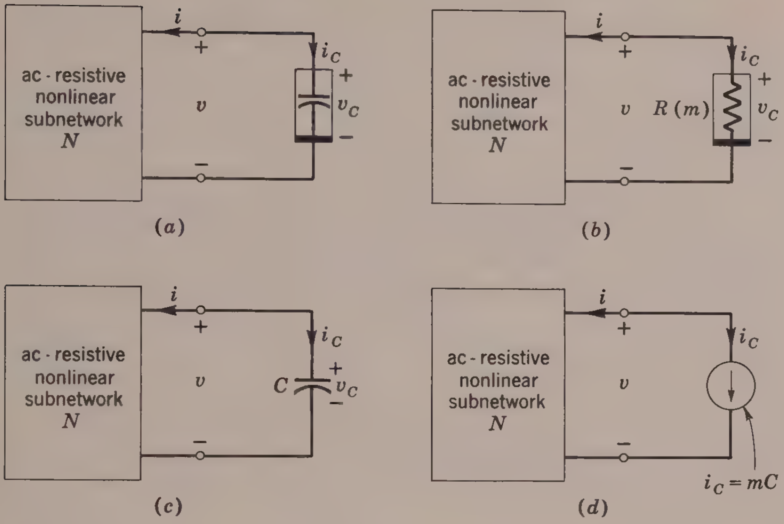
as shown in Fig. 18-8 are obtained. It is interesting to observe that for this particular vector field, the slope of the directors along the $m = 1$ constant-slope curve $v_o = t - 1$ is also equal to 1, and hence all $m = 1$ directors must lie along the line $v_o = t - 1$. This means that no solution curve can cross the $m = 1$ constant-slope curve, for as soon as any solution curve intersects this line, it must follow the line from then on. If the solution curve is smooth, which is the case here, it will not intersect this line abruptly, and hence we conclude that the complete response of this network corresponding to any initial state is asymptotic to the line $v_o = t - 1$. Such an asymptote is often called a *separatrix* because it separates the solution curves with initial states on each side of the asymptote.

The above observation is one example of the qualitative information which is often much more useful to engineers than the finding of one solution curve. In addition to the qualitative information that the present approach can provide, observe also that once the vector field is obtained as in Fig. 18-8, it is a trivial matter to sketch the solution curve corresponding to any initial state $v_o(t_0)$, including the zero-state response which corresponds to $v_o(t_0) = 0$. The zero-state response for this network had been obtained earlier analytically as a function of the time constant τ (Eq. 18-6). Hence, as a check the zero-state response corresponding to $\tau = 1$ is given by

$$v_o(t) = t - (1 - e^{-t}) \qquad t \geq 0$$

The graph of this waveform is easily seen to be identical with the solution curve 1 in Fig. 18-8, as it should.

In order to apply the constant-slope method, it is necessary to derive first the normal-form equation before we can obtain Eq. (18-37) for plotting the constant-slope curves. No difficulty was encountered in the above example because the network was linear. Unfortunately, as we have indicated many times in the past, it is often difficult, if not impractical, to obtain the normal-form equation of a nonlinear network in analytic form. Hence, in order for this powerful method to be of much practical value, we must find an easier method for drawing the constant-slope curves without having to derive the normal-form equation first. The key to this method lies in the definition of a *constant-slope network*.



(a) (b) (c) (d)

Fig. 18-9. A nonlinear (linear) capacitor-loaded first-order network and its associated constant-slope network.

18-4-2 THE CONCEPT OF CONSTANT-SLOPE NETWORKS

Consider the first-order capacitor-loaded network shown in Fig. 18-9*a*. For simplicity let us assume that the nonlinear capacitor is characterized by a *voltage-controlled v-q* curve.[1] The equation of motion relating the terminal variables is given by

$$i_C = \frac{dq}{dt} = \frac{dq}{dv_C}\frac{dv_C}{dt}$$

or

$$\frac{dv_C}{dt} = \frac{i_C}{C(v_C)} \tag{18-40}$$

where

$$C(v_C) \equiv \frac{dq(v_C)}{dv_C} \tag{18-41}$$

is the *incremental capacitance,* which is a *single-valued function* of v_C because the capacitor is assumed to be voltage-controlled. In view of the uniqueness hypothesis, the normal-form equation of this network can be written in the form of Eq. (18-35).[2]

$$\frac{dv_C}{dt} = f(v_C,t) \tag{18-42}$$

[1] A similar procedure can be developed if the capacitor is *charge-controlled.* See Prob. 18-3.

[2] Observe that in order for $f(v_C,t)$ to be a single-valued function, it is necessary that the DP plot of the resistive subnetwork N in Fig. 18-9*a* be *voltage-controlled* for all time t.

Since both Eqs. (18-40) and (18-42) refer to the same network, we must have

$$f(v_C,t) = \frac{i_C}{C(v_C)} \tag{18-43}$$

The constant-slope curve defined by Eq. (18-37) is obtained by equating Eq. (18-43) to m and plotting the locus of all points on the t-v_C plane which satisfy this relationship;[1] namely,

$$\frac{i_C}{C(v_C)} = m \tag{18-44}$$

If we rewrite Eq. (18-44) into the form

$$i_C = mC(v_C) \tag{18-45}$$

we can interpret this relationship to be a v_C-i_C curve of a nonlinear resistor. Hence, the constraint that must be satisfied by the constant-slope curve can be "simulated" by connecting a nonlinear resistor $R(m)$ with the v_C-i_C curve defined by Eq. (18-45) in place of the capacitor as shown in Fig. 18-9*b*.[2] We shall call the resulting resistive network in Fig. 18-9*b* the *constant-slope network*. Since any point on the constant-slope curve must satisfy Eq. (18-44), and since the constraint imposed by Eq. (18-44) is equivalent to that of replacing the nonlinear network by its associated constant-slope network, it follows that the solution waveform $v_C(t,m)$ across $R(m)$ is the constant-slope curve corresponding to slope m.

Since the constant-slope network is a resistive network, the solution $v_C(t,m)$ can be obtained by the methods developed in Part 2 of this book or by numerical techniques. *The important point to realize is that we have transformed a dynamic network problem into a resistive network problem.* The v_C-i_C curve of the nonlinear resistor $R(m)$ is obtained by simply multiplying the incremental capacitance $C(v_C)$ by the desired slope m. In the special, but rather common, case where the capacitor is linear as shown in Fig. 18-9*c*, the incremental capacitance $C(v_C) = C$, its capacitance, and the v_C-i_C curve becomes the horizontal line $i_C = mC$. In other words, the nonlinear resistor $R(m)$ reduces to an independent dc current source as shown in Fig. 18-9*d*. The above results are of fundamental importance, and we shall summarize them in the form of a theorem.

---

[1] Observe that both i_C and v_C are functions of time.

[2] Since corresponding to each value m, a different v_C-i_C curve must be used, we have chosen the symbol $R(m)$ to denote its associated slope m. Similarly, we shall denote the corresponding voltage $v_C(t)$ by $v_C(t,m)$.

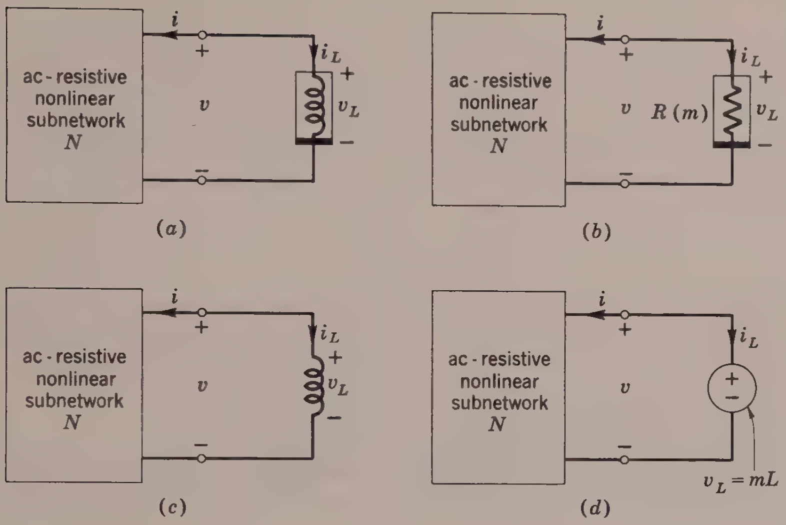


Fig. 18-10. A nonlinear (linear) inductor-loaded first-order network and its associated constant-slope network.

CONSTANT-SLOPE CAPACITOR-LOADED NETWORK THEOREM[1]

The constant-slope curve of any nonautonomous, voltage-controlled, capacitor-loaded, first-order network corresponding to any slope m can be determined by solving for the voltage $v_C(t,m)$ across the nonlinear resistor $R(m)$ of the associated constant-slope network. The v_C-i_C curve of $R(m)$ is given by $i_C = mC(v_C)$, where $C(v_C)$ is the incremental capacitance of the nonlinear capacitor. If the capacitor is linear, with capacitance C, the nonlinear resistor $R(m)$ reduces to an independent dc current source with terminal current $i_C = mC$.

By the principle of duality we obtain the corresponding constant-slope network of a current-controlled, inductor-loaded, first-order network as shown in Fig. 18-10. The dual theorem can now be stated as follows:

CONSTANT-SLOPE INDUCTOR-LOADED NETWORK THEOREM

The constant-slope curve of any nonautonomous, current-controlled, inductor-loaded, first-order network corresponding to any slope m can be determined by solving for the current $i_L(t,m)$ through the nonlinear resistor $R(m)$ of the associated constant-slope network. The v_L-i_L curve of $R(m)$ is given by $v_L = mL(i_L)$, where $L(i_L)$ is the incremental inductance of the nonlinear inductor. If the inductor is linear with inductance L, the nonlinear resistor $R(m)$

---

[1] For a discussion of some related results, see L. O. Chua, Two New Theorems in Nonlinear Networks, *Midwest Symp. Circuit Theory, 11th, University of Notre Dame, South Bend, Ind.*, pp. 340–351, May 13–14, 1968.

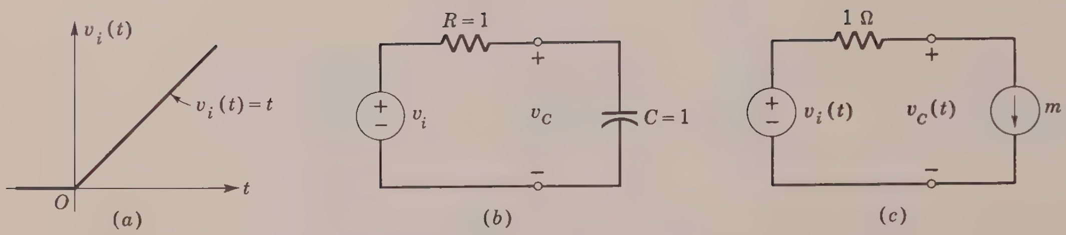


Fig. 18-11. A linear *R-C* network and its associated constant-slope network.

reduces to an independent dc voltage source with terminal voltage $v_L = mL$.

18-4-3 SOME ILLUSTRATIVE EXAMPLES

Let us now apply the constant-slope network approach to obtain the complete response of some practical networks.

EXAMPLES

1. Consider the linear *R-C* network and its unit-ramp voltage source shown in Fig. 18-11.[1] Since the capacitor is linear with $C = 1$, the constant-slope network is simply obtained as shown in Fig. 18-11*c*. The constant-slope curve corresponding to slope m is identical with the voltage waveform $v_C(t,m)$, which in this trivial case is given by

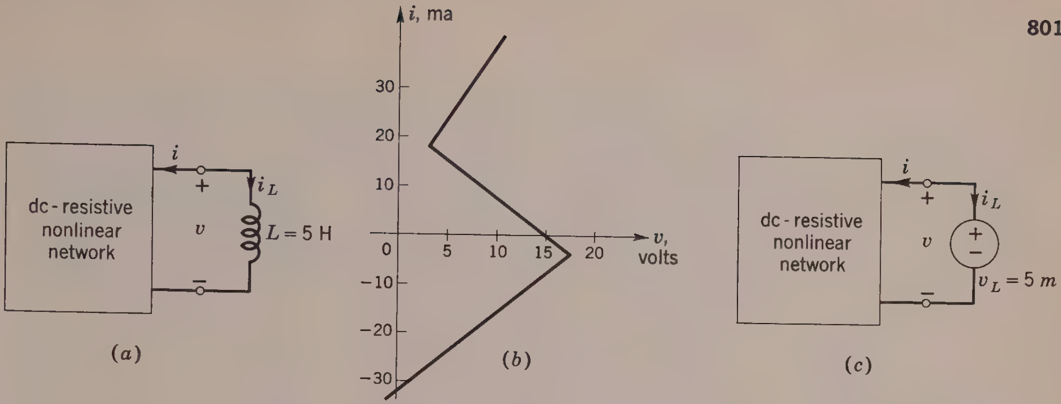
   $$v_C(t,m) = v_i(t) - m(1) = t - m \qquad (18\text{-}46)$$

   Observe that Eq. (18-46) is identical with Eq. (18-39), and hence the same solutions shown in Fig. 18-8 will be obtained.

2. In order to demonstrate that the constant-slope approach is sometimes preferable even in the case of autonomous networks, let us consider the first-order, autonomous, inductor-loaded network shown in Fig. 18-12*a*. In order to compare this method with the piecewise-linear approach developed in Chap. 14, the DP plot shown in Fig. 18-12*b* is chosen similar to that shown in Fig. 14-22*b* (see Example 2 in Sec. 14-8-1). The constant-slope network for this example is shown in Fig. 18-12*c*. The constant-slope curve corresponding to each value of m is obtained by finding the current $i_L(t)$ corresponding to the terminal voltage $v_L = m(5)$. Since the present network is autonomous, $i_L(t,m)$ is simply the *negative* of the operating-point current i. Based on this observation, we can conclude that *the constant-slope curves of any autonomous network are always horizontal lines.*

---

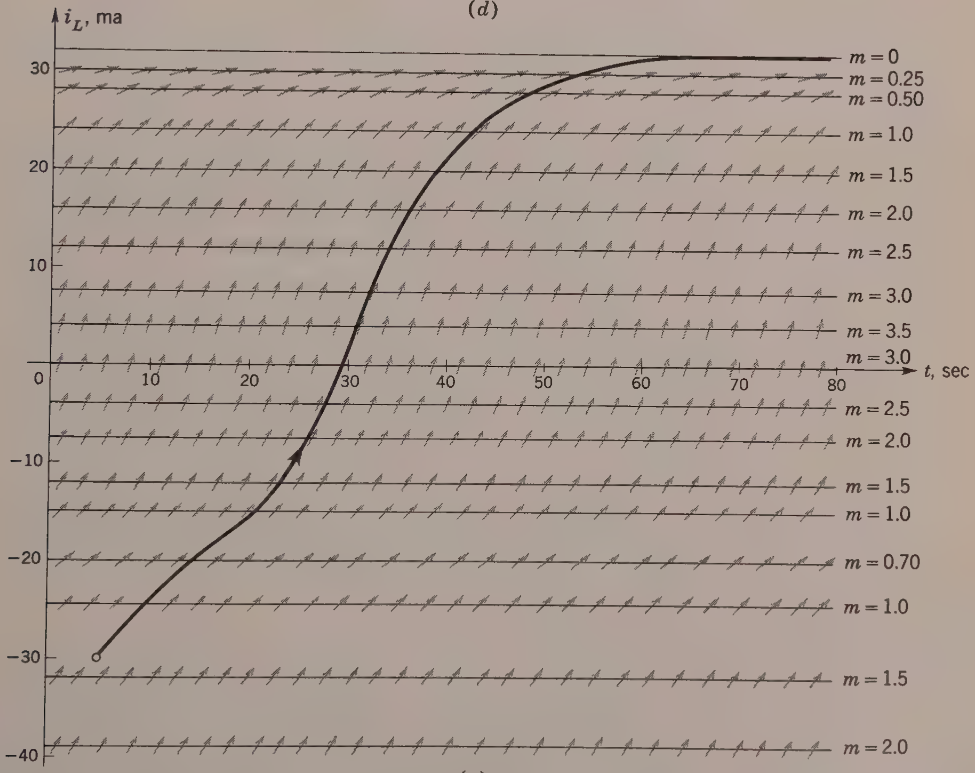
[1] This example was considered earlier in Sec. 18-4-1.



(a) (b) (c)

m	0.0	0.25	0.50	0.70	1.0	1.5	2.0	2.5	3.0	3.5
v_L	0.0	1.25	2.50	3.50	5.0	7.5	10.0	12.5	15.0	17.5
i_L	32	30.0	28.0	26.0	24.0	20.0	16.0	12.0	7.5	4
				−17.0	−15.0	−12.0	−7.5	−4.0	0	
				−20.0	−24.5	−32.0	−39.0			

(d)



(e)

Fig. 18-12. An example illustrating the analysis of an autonomous first-order network by the constant-slope network approach.

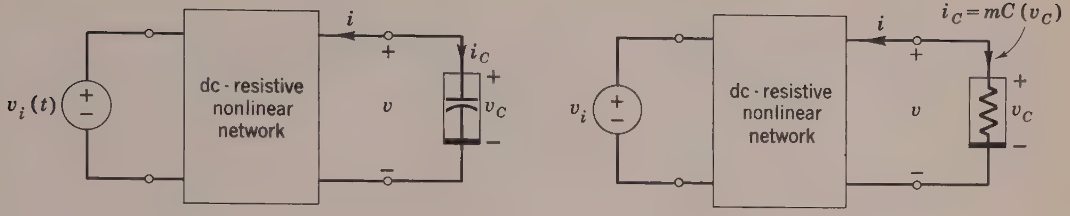
Since the DP plot is not monotonic, there will be more than one solution $i_L(t,m)$ corresponding to a certain range of values of m. For ease of reference, the solutions $i_L(t,m)$ corresponding to 10 typical values of m are tabulated in Fig. 18-12*d*. The constant-slope curves corresponding to some values of m consist of three separate branches. The vector field associated with these constant-slope curves (horizontal lines in this example) is easily constructed from the data listed in Fig. 18-12*d* and is shown in Fig. 18-12*e*. The solution waveform (zero-input response) $i_L(t)$ corresponding to any initial state $i_L(t_0)$ can now be trivially sketched.[1] For example, the solution $i_L(t)$ corresponding to the initial state $i_L(5) = -30$ ma is shown in Fig. 18-12*e*. Observe that this waveform agrees with that obtained earlier in Fig. 14-26*a*, as it should.

The obvious question that one raises at this point is, why bother with so many methods? The answer is that each method has some desirable features which may not be present in the others. For example, the constant-slope method has the obvious advantage that one could obtain the solution corresponding to *any number* of initial states without much additional work. In contrast with this, the piecewise-linear approach developed in Chap. 14 would require a separate analysis for each additional initial state. On the other hand, the constant-slope method suffers from the disadvantage that once an element parameter, such as the time constant, is changed, a new vector field must be obtained. In short, the constant-slope method is preferred if the behavior of the solution waveforms with respect to changes in the initial states is desired, whereas the piecewise-linear method is preferred if the behavior of the solution waveform with respect to changes in the element parameters is desired. Another advantage of the constant-slope method is that the DP plot need not be in piecewise-linear form. Since the amount of work required by the piecewise-linear approach increases with the number of segments, it is clear that the constant-slope method is much easier to apply in cases where the DP plot is highly nonlinear.

3. Consider next the class of nonautonomous networks shown in Fig. 18-13*a* where there is only one ac voltage source, and where the capacitor is nonlinear. This class of nonlinear networks is commonly encountered in integrated circuits where the nonlinear capacitor usually arises because of the nonlinear capacitance between semiconductor junctions. The constant-

---

[1] The actual solution $i_L(t)$ is not to be confused with the constant-slope solution $i_L(t,m)$. While $i_L(t)$ is a unique physical solution, the solution $i_L(t,m)$ corresponding to each value m is a fictitious nonphysical solution and is generally a multivalued function of time, possibly with separate branches.



(*a*) (*b*)

Fig. 18-13. An important class of nonautonomous first-order networks consists of a single ac voltage source, a dc-resistive nonlinear network, and a nonlinear capacitor.

slope network is shown in Fig. 18-13*b*, and the constant-slope curve can be found from the solutions $v_C(t,m)$. The most general method for finding $v_C(t,m)$ is to derive the TC plot between v_C and v_i corresponding to each value of *m*. If we denote this TC plot by[1]

$$v_C = T(v_i,m) \qquad (18\text{-}47)$$

then the solution $v_C(t,m)$ is simply obtained by a graphical composition between this TC plot and the input voltage waveform $v_i(t)$.

4. As a final example, consider the capacitor-loaded transistor amplifier circuit shown in Fig. 18-14*a*. The corresponding constant-slope network is shown in Fig. 18-14*b*. Ideally, we would like to obtain an amplified version of the input signal $v_i(t)$. However, because of the nonlinearity of the transistor as well as the presence of the capacitor, the output waveform will not be a faithful reproduction. The amount of *distortion*, i.e., deviation from the ideal response, has been observed experimentally to increase as the amplitude of the input signal increases. In order to obtain a good qualitative picture of the nature of the distortion, let us consider the three typical input waveforms shown in Fig. 18-14*c* to *e*.

   Notice that the network in Fig. 18-14*a* is a member of the class of networks shown in Fig. 18-13*a*. Hence, the systematic method for obtaining the solution waveforms $v_C(t,m)$ is to derive the TC plot $v_C = T(v_i,m)$ of the constant-slope network shown in Fig. 18-14*b*. This has been done graphically, and the resulting TC plots corresponding to several values of *m* are shown in Fig. 18-15. It remains to obtain a graphical composition between the three given input waveforms with each of these TC plots to obtain the corresponding vector fields shown in Fig. 18-16*a* to *c*. The complete response corresponding to each of the three arbitrarily chosen initial states is plotted on

---

[1] In general this TC plot will be multivalued and may contain separate branches.

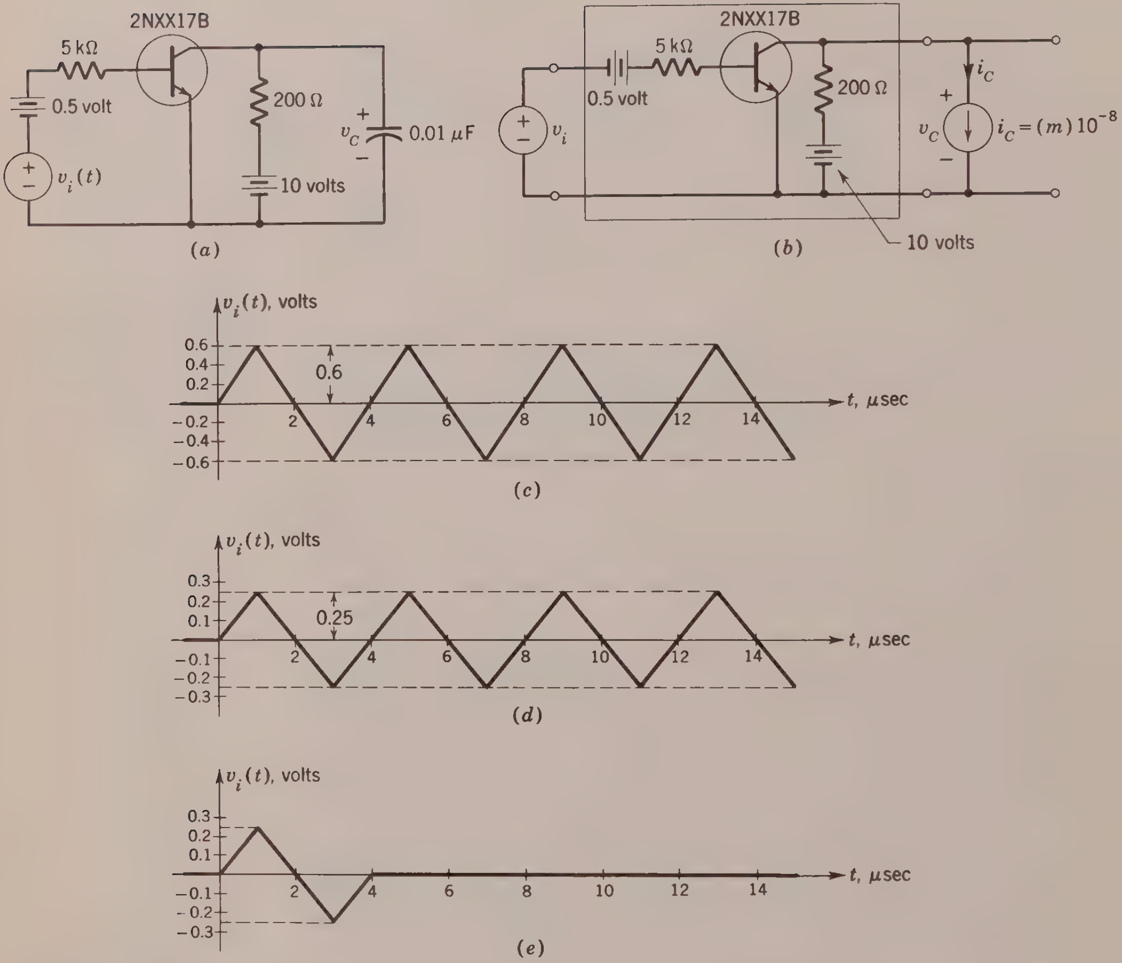


Fig. 18-14. A capacitor-loaded transistor amplifier circuit.

top of these vector fields. This example was carried out graphically for pedagogical reasons. In practice, it would be much easier to write a computer program, such as the Meca program described in Sec. 12-5-2, to construct the constant-slope curves directly.

An examination of these waveforms shows that the output waveforms suffer from two types of distortion. First, there is the *flattening* effect due to the nonlinearities of the TC plot for large positive voltages. We shall refer to this type of distortion as the *nonlinear distortion.* Second, there is a *smoothing* effect due to the presence of the linear capacitor. We shall call this the *linear distortion.* From our previous experience in Chaps. 14 and 15 we know that the linear distortion can be minimized by reducing the time constant of the network. The nonlinear distortion is harder to get rid of because it is inherent in the transistor characteristics. One of the most important problems in the design of a high-fidelity amplifier is to minimize both the nonlinear and the linear distortions.

Exercise 1: Sketch the vector field associated with the normal-form equation

$$\frac{di}{dt} = i + t^3 \qquad t \geq 0$$

Exercise 2: Formulate some simple rules for choosing an appropriate time scale for drawing the constant-slope curves so that the slope of the directors corresponding to different values of m can be conveniently found.

Exercise 3: Obtain the vector field of the network shown in Fig. 18-12*a* with $L = 10$ mH.

Exercise 4: Using the 2NXX17B transistor curves given in Appendix D, obtain the TC plots shown in Fig. 18-15.

Exercise 5: Construct the vector field shown in Fig. 18-16.

Exercise 6: Using the TC plots shown in Fig. 18-15, find the output waveform corresponding to a 1-volt peak-to-peak 500-kHz sine wave and an initial state $v_C(0) = 4$ volts.

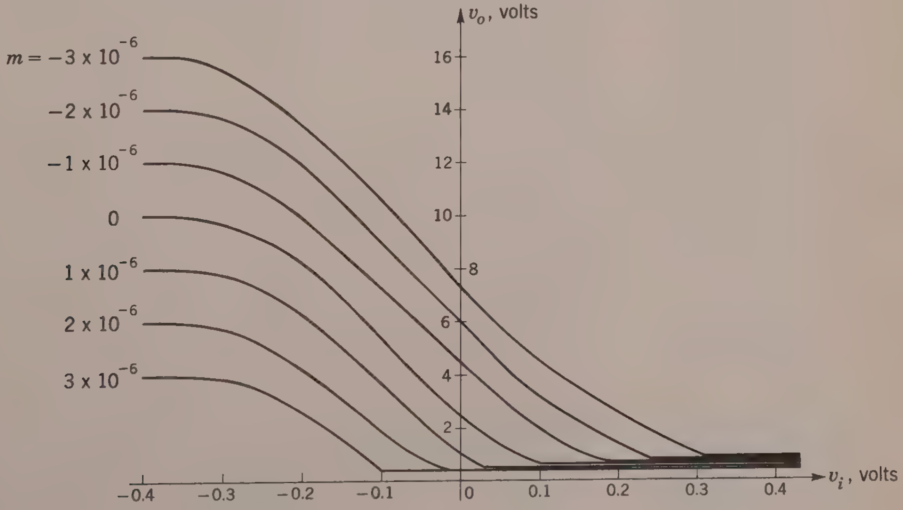


Fig. 18-15. The TC plots $v_o = T(v_i, m)$ associated with the constant-slope network of the capacitor-loaded amplifier circuit.

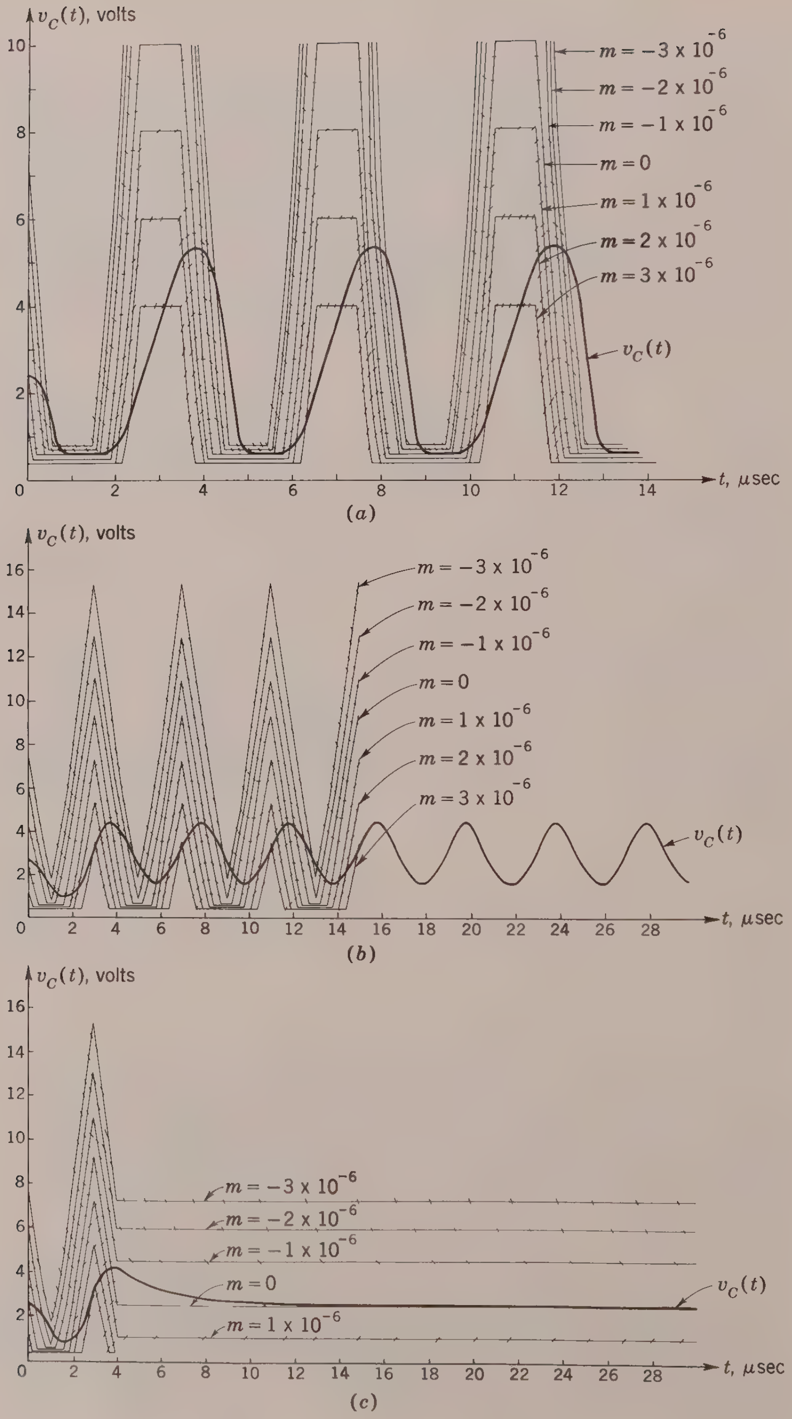


Fig. 18-16. The output waveforms of the capacitor-loaded transistor amplifier circuit.

18-5 SOME PRACTICAL APPLICATIONS OF NONAUTONOMOUS FIRST-ORDER NONLINEAR NETWORKS

Now that we have learned a general method for analyzing any nonautonomous first-order nonlinear network, let us turn to the *synthesis* of some practical circuits. In line with our basic philosophy of not studying every individual circuit, as is often done in the traditional "cookbook" approach, we shall try merely to extract some basic concepts which have been found useful in the synthesis of this class of networks.

Most nonautonomous, first-order, nonlinear networks are designed to transform or process an input signal $v_i(t)$ into an output signal $v_o(t)$ with some prescribed properties. We have already encountered this practical problem in Chaps. 9 and 10, and indeed we have shown there that any transformation property which can be translated into the form of a TC plot can be realized by a resistive nonlinear network. Hence, our concern here will be restricted to those specifications which require a dynamic network realization, such as the integrator and the differentiator.

Most applications which call for a nonautonomous, first-order, nonlinear network are based on three useful properties of some networks containing ideal diodes,[1] namely, *the monotonic charging property, the pulse-smoothing property,* and *the pulse-sharpening property.*

18-5-1 THE MONOTONIC CHARGING PROPERTY AND ITS APPLICATIONS

Consider the network shown in Fig. 18-17*a* where the black box N represents an arbitrary resistive or dynamic subnetwork.[2] Since the ideal diode is in series with the capacitor, $i_d = i_C = C\, dv_C/dt$. But $i_d \geq 0$ for an ideal diode; hence, we conclude that $dv_C/dt \geq 0$, independent of the waveform $v_i(t)$ across the black box N. This means that the voltage across the capacitor is a nondecreasing function of time. Physically, this property implies that the capacitor cannot discharge. In order to obtain a relationship between $v_i(t)$ and $v_C(t)$, let us observe that the ideal diode is an open circuit whenever $v_d = v_i(t) - v_C(t) \leq 0$. It follows, therefore, that the capacitor voltage will remain unchanged ($dv_C/dt = 0$) whenever $v_i(t) \leq v_C(t)$. If $v_i(t)$ attains its maximum value $E_{\max}$ at some time $t = t_{\max}$, it follows that $v_C(t) = E_{\max}$ for all times $t \geq t_{\max}$. Hence, the capacitor is equivalent to a battery with terminal voltage $E_{\max}$ for all times $t \geq t_{\max}$, as shown in Fig. 18-17*b*. If we transpose the diode terminals, and if $v_i(t)$ attains its minimum value $-E_{\min}$ at

---

[1] The choice of an ideal diode greatly simplifies matters. Since ideal diodes can be closely approximated by junction diodes, we expect that the actual behavior of the network will not differ considerably from the ideal case (see the small-variation postulate in Sec. 1-6-5). This approach to the synthesis problem is extremely useful in practice and is often the key for arriving at an initial design. Any modification and refinement can then be added later, and an exact analysis carried out by the general methods developed earlier.

[2] For simplicity, N is usually replaced by an equivalent voltage source.

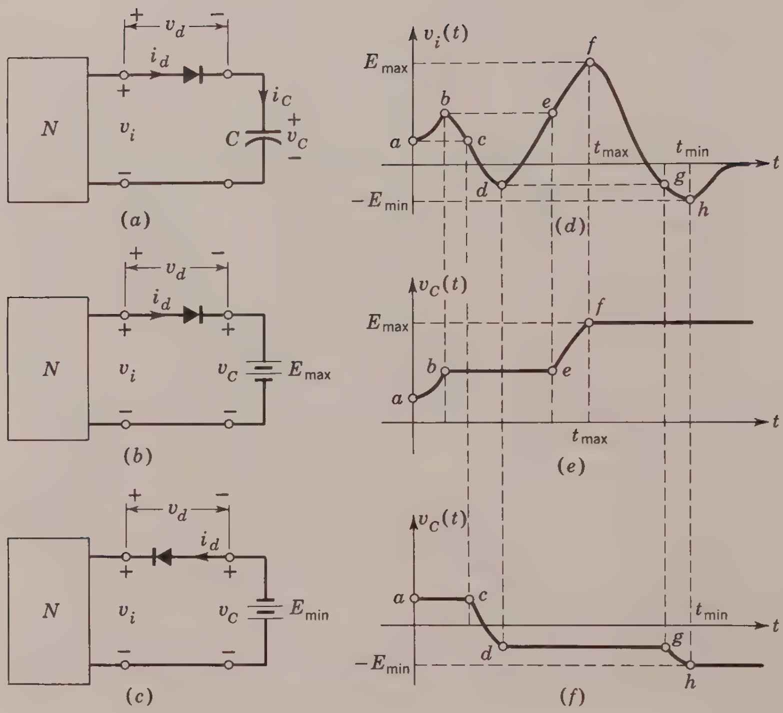


Fig. 18-17. An illustration of the monotonic charging property of a capacitor–ideal diode network.

$t = t_{\text{min}}$, then a similar argument to the one above will show that the network is equivalent to that shown in Fig. 18-17*c* for all times $t \geq t_{\text{min}}$. Let us summarize the above results as follows:

MONOTONIC CHARGING PROPERTY

The capacitor voltage $v_C(t)$ of the network shown in Fig. 18-17*a* will follow the waveform $v_i(t)$ in such a way that it never decreases with time. This implies that $v_C(t)$ will assume the constant value $v_C(t_j)$ for all times $t \geq t_j$ such that $v_i(t) \leq v_C(t_j)$. Moreover, if $t = t_{\text{max}}$ is the time where $v_i(t)$ attains its maximum value E_{max}, then $v_C(t) = E_{\text{max}}$ for all times $t \geq t_{\text{max}}$, and is said to be in steady state. For purposes of analysis, the capacitor can be replaced by a battery with terminal voltage E_{max} during steady state, as shown in Fig. 18-17*b*.

As an application of this property, let us find the waveform $v_C(t)$ of the network in Fig. 18-17*a* corresponding to the input waveform $v_i(t)$ shown in Fig. 18-17*d*. Starting at point *a*, $v_C(t) =$

$v_i(t)$ for all times up to point b when $v_i(t)$ begins to decrease. Since $v_C(t)$ cannot decrease, it must remain constant between points b and e. It will follow $v_i(t)$ again between points e and f. Since $v_i(t)$ attains its maximum value at point f, the capacitor has reached its steady state, and $v_C(t) = E_{\max}$ from then on. The complete waveform for $v_C(t)$ is shown in Fig. 18-17*e*.

If the diode terminals are transposed, then by a similar analysis we arrive at the waveform $v_C(t)$ shown in Fig. 18-17*f*. Observe that in this case, the steady state occurs at $t = t_{\min}$, and the capacitor can be replaced by the battery with terminal voltage $E_{\min}$ as shown in Fig. 18-17*c*.

The above capacitor-battery substitution property has been applied in the design of many practical circuits. We shall consider three common applications.

Clamping circuits A very common requirement in the design of electronic circuits is to change the dc level of an input periodic waveform $v_i(t)$ so that either the maximum or the minimum value of $v_i(t)$ is fixed at some prescribed value. An examination of the equivalent circuits shown in Fig. 18-17*b* and *c* reveals that the voltage v_d across the ideal diode is equal to $v_i - E_{\max}$ and $v_i + E_{\min}$, respectively, in steady state. Hence, if we redraw these networks as shown in Fig. 18-18*a* and *b*, we would obtain the sequence of waveforms shown in Fig. 18-18*c* and *d*, respectively. Observe that the maximum value of the output waveform $v_o(t)$ in Fig. 18-18*c* is "clamped" at ground level. In contrast with this, the minimum value is being clamped at ground level in Fig. 18-18*d*. In view of the above interpretations, these networks are called *clamping networks* (also known as dc restorers). If it is desired to clamp the input waveform at other than ground level, a battery with an appropriate voltage and polarity can be inserted in series with the ideal diode. But why not just connect a battery in series with the voltage source, thereby removing the capacitor and diode? This is all right if the voltage amplitude of $v_i(t)$ is fixed. However, if the amplitude changes, as it normally does in practice, then the battery voltage must also change in order to maintain the same clamping level. The clamping circuit in Fig. 18-18 does this automatically and may, therefore, be interpreted as a variable-voltage battery.

What happens when we replace the ideal diode in Fig. 18-18 with a junction diode? Obviously, the monotonic charging property is no longer valid. An exact analysis for this case can be easily made by using the constant-slope approach developed in Sec. 18-4.

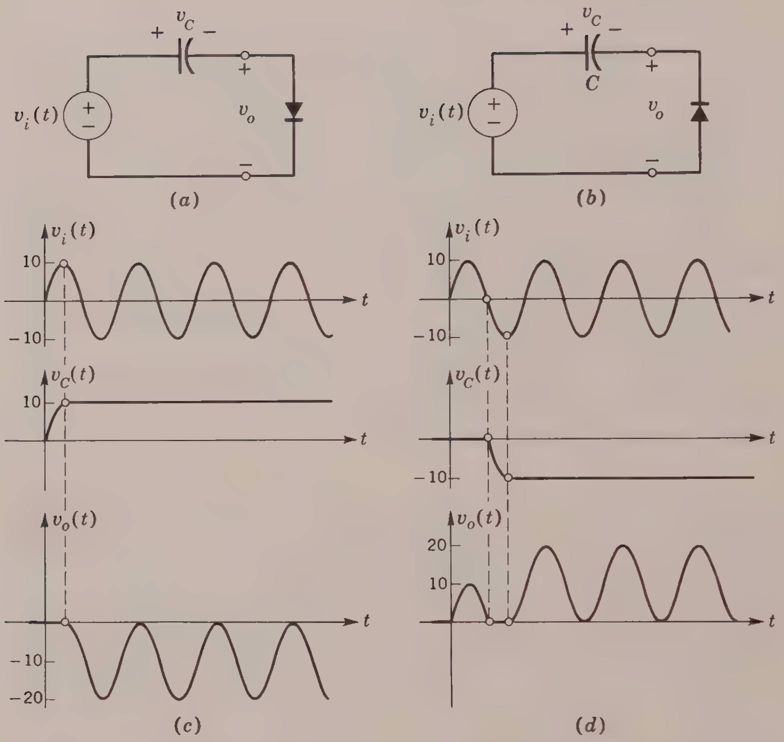


Fig. 18-18. The waveforms associated with two typical clamping networks.

Filtered rectifier An examination of the waveforms in Fig. 18-18 shows that $v_C(t)$ is a dc voltage in steady state. Here is a simple means for transforming a sinusoidal voltage into a dc voltage of the same amplitude. In other words, if we take our output voltage across the capacitor, the circuit functions as a *filtered rectifier.* In practice we cannot obtain the perfect dc voltage shown in Fig. 18-18 for two reasons. First, the ideal diode can only be approximately realized in practice by the junction diode, and hence the monotonic charging property is only approximately satisfied. Second, the output voltage $v_C(t)$ is invariably used to drive other circuits, and, hence, to be realistic, a load resistor R_L must be connected across the capacitor, as shown in Fig. 18-19*a*. This *loading effect* will seriously alter the ideal filtered rectifier waveform. An exact analysis of this circuit can be carried out easily by the constant-slope approach. A typical output waveform for this rectifier is shown in Fig. 18-19*b*.

Voltage doubler Even in the ideal case the magnitude of the dc voltage of a filtered rectifier cannot exceed the amplitude of the

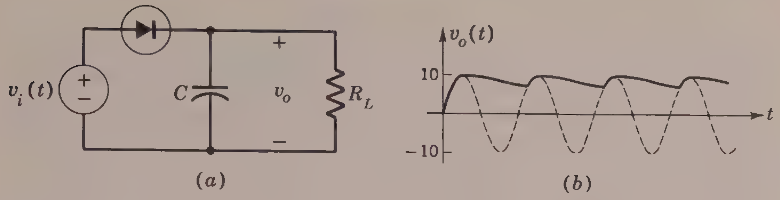


Fig. 18-19. A filtered rectifier and its output waveform.

sinusoidal input voltage. In many practical applications higher dc voltages are often required. One simple, but rather expensive, method is to connect two or more filtered rectifiers in series. A more practical circuit can be designed by observing that the amplitude of $v_o(t)$ of the clamping network in Fig. 18-18*d* is twice the magnitude of the input waveform. Hence, if we connect the output terminals of the clamping network in Fig. 18-18*b* in place of the voltage source in the filtered-rectifier circuit in Fig. 18-19, we would obtain the *voltage-doubler circuit* shown in Fig. 18-20. If we assume the ideal case where R_L is an open circuit, then we can apply the monotonic charging property and conclude that the voltage $v_o(t)$ across the capacitor of the filter rectifier circuit is a nondecreasing function of time. Moreover, in steady state no current flows into this capacitor, and the clamping network will not have any loading effect. Hence, the output voltage magnitude across the clamping network in steady state is equal to twice the input magnitude as shown in Fig. 18-18*d*. It follows, therefore, that the output voltage $v_o(t)$ of the filtered rectifier in steady state is also equal to twice the input voltage. This accounts for the name voltage doubler. In practice, the load resistor R_L is not infinite, and the junction diodes are not ideal. In this case, the network must be analyzed as a second-order nonautonomous network.

18-5-2 THE PULSE-SMOOTHING PROPERTY

The monotonic charging property is never exactly realized in practice, not only because of the ideal diode, but also because the current drawn by the external load connected across the capacitor is

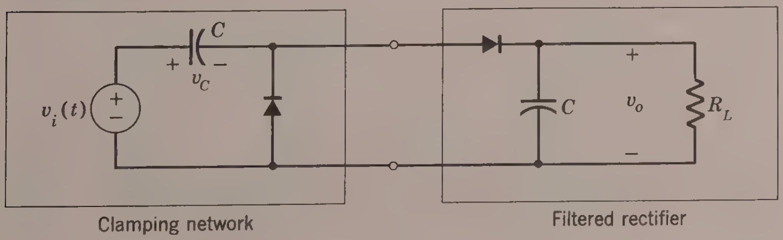
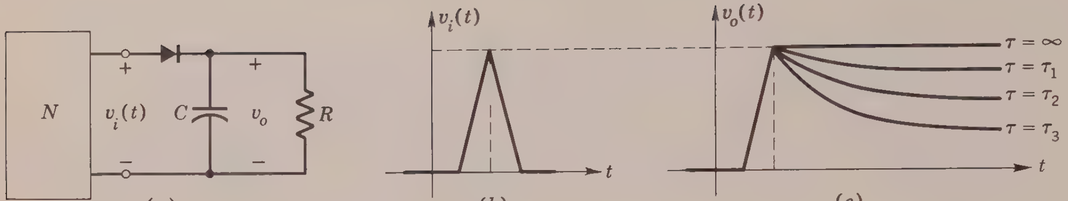


Fig. 18-20. A practical voltage-doubler network can be interpreted as the cascade connection between a clamping network and a filtered rectifier.



(a) (b) (c)

Fig. 18-21. An illustration of the pulse-smoothing property due to a finite time constant introduced by the load resistor R.

never exactly zero. To investigate this loading effect, let us redraw the network in Fig. 18-17*a* as shown in Fig. 18-21*a*. As before, the black box N represents an arbitrary network, usually substituted by an equivalent voltage source. Since the ideal diode becomes an open circuit whenever $v_i(t) < v_o(t)$, the capacitor voltage will "discharge" through the resistor R with a time constant $\tau = RC$. This corresponds to the zero-input response of the parallel RC network. To investigate the effect of the time constant, consider the single pulse waveform $v_i(t)$ shown in Fig. 18-21*b*. The output waveforms corresponding to $\tau = \infty$, τ_1, τ_2, $\tau_3(\tau_1 > \tau_2 > \tau_3)$ are shown in Fig. 18-21*c*. Observe that as the time constant decreases, the waveform deteriorates from the ideal clamping output ($\tau = \infty$). While this deterioration is undesirable for a clamping circuit, it is useful for other applications where it is desirable to smooth out a sharp pulse. Let us consider some practical applications based on this pulse-smoothing property.

Pulse stretchers In the design of many instrumentation circuits, it is often required to transform a voltage pulse, e.g., the output pulse from a monostable multivibrator, by "stretching" it to obtain an expanded pulse. By an appropriate choice of the time constant τ, the circuit shown in Fig. 18-21*a* will function as a pulse stretcher. Notice that this circuit is identical with the filtered-rectifier circuit shown in Fig. 18-19*a*. The only difference here is that the input voltage $v_i(t)$ of the rectifier is usually a periodic sine wave, whereas the input waveform of a pulse stretcher is usually a single pulse. In the case of the rectifier, the time constant should be chosen as large as possible, whereas for the pulse stretcher the time constant depends on how much stretching is desired.

An *exact* analysis of the filtered-rectifier or the pulse-stretcher circuit can be obtained by the constant-slope network method. However, by assuming the diode to be ideal, it is possible for us to obtain an approximate analysis which is sometimes more illuminating.[1] For example, if we replace the junction diode in the filtered-rectifier circuit of Fig. 18-19*a* by an ideal diode, then we can immediately obtain the output waveform $v_o(t)$ as shown by the solid curve in Fig. 18-22 (the corresponding sinusoidal input waveform is shown dotted). Comparing this approximate solution with the exact solution shown in Fig. 18-19*b* shows that they differ only very slightly; the slope of the exact waveform is continuous, whereas the slope of the approximate waveform has two discontinuities per cycle. Since this waveform is periodic for all times $t \geq T/4$, it suffices to consider only one cycle of the periodic waveform $(T/4,\ T + T/4)$. Observe that the waveform approaches a true dc voltage as $\tau \to \infty$. The departure from this ideal case is measured by the amount of *ripple*

---

[1] One should not insist on using a particular method because it is the most general. In fact, the price one pays for generality is usually increased complexity. There are many practical circuits which can be analyzed by simpler shortcuts. The only caution here is that one must always precisely understand the restrictions and simplifying assumptions that are invariably associated with any shortcuts.

$$\Delta E \equiv E - E_1 = E[1 - e^{-(T_1 - T/4)/\tau}] \tag{18-48}$$

To function as a filtered rectifier, the time constant τ must be chosen as large as possible. Under this condition we can approximate $T_1 - T/4 \approx T$, and the ripple reduces to

$$\Delta E = E(1 - e^{-T/\tau}) \tag{18-49}$$

The amount of ripple ΔE that one can tolerate depends, of course, on how large ΔE is with respect to the amplitude E. Accordingly, the filtered rectifier is often rated in practice by the *per-unit ripple* $\eta \equiv |\Delta E/E|$. If ω is the frequency of the sinusoid, then $T/\tau = 2\pi/\omega\, RC$, and the per-unit ripple will be small if $\omega RC \gg 1$. Under

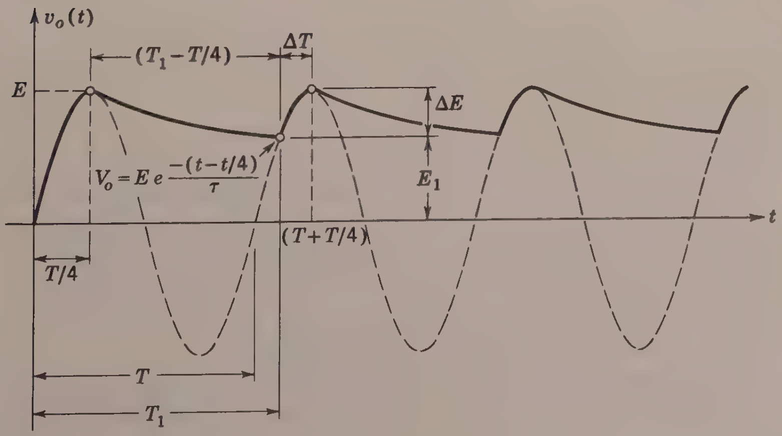


Fig. 18-22. The approximate output waveform of a filtered rectifier.

this condition, $T/\tau \ll 1$, and the per-unit ripple can be approximated by[1]

$$\eta = |1 - e^{-T/\tau}| \approx \frac{T}{RC}$$

This relationship is valid, of course, only if the time constant τ is large compared with the period T of the input sinusoid.

AM envelope detector The signal transmitted by an AM radio station is in the form of an *amplitude-modulated* wave[2]

$$v(t) = [E + mE\, v_s(t)] \sin \omega_c t \tag{18-50}$$

where $v_s(t)$ is the message- or information-bearing signal that is being transmitted. The term $\sin \omega_c t$ represents a very high-frequency sinusoid, where ω_c is called the *carrier* frequency. The parameter m is a number less than unity and is called the *modulation index*. Since Eq. (18-50) is the signal that is being received, an essential task of any AM radio is to recover the information-bearing signal $v_s(t)$. Any circuit for accomplishing this task is called an AM *demodulator* or *detector*. We shall now show that, through an appropriate choice of time constants, the pulse-stretcher circuit shown in Fig. 18-21*a* can be designed to function as a demodulator.

To be specific, let us suppose that the information-bearing signal consists of a cosine waveform

$$v_s(t) = E \cos \omega_s t \tag{18-51}$$

where $\omega_s \ll \omega_c$. The AM signal is, therefore, given by

$$v(t) = E(1 + m \cos \omega_s t) \sin \omega_c t \tag{18-52}$$

as shown in Fig. 18-23*a*. Observe that the signal $v_s(t)$ that we want to recover in this case forms the *envelope* (shown dotted) of this waveform. Accordingly, the detector circuit for reproducing the information-bearing signal is called an *envelope detector*. Our strategy will be to recover first the envelope waveform. The desired signal differs only by a dc level and can be easily obtained by using the clamping network developed earlier.

One method for recovering the envelope of the AM waveform in Fig. 18-23*a* consists of *stretching* the positive peaks of this

---

[1] Recall the well-known expansion

$$e^x = 1 + x + \frac{1}{2}x^2 + \frac{1}{3!}x^3 + \frac{1}{4!}x^4 + \cdots \approx 1 + x \qquad x \ll 1$$

[2] In order to accomplish long-distance wireless transmission, it is necessary to modulate a very high-frequency carrier by the signal $v_s(t)$. The amplitude modulation is one common form of modulation.

signal by the pulse-stretcher circuit shown in Fig. 18-21*a*. However, in order to obtain as good an approximation as possible, the time constant $\tau = RC$ must be carefully chosen. If the time constant is too small, the output waveform will contain too much ripple as shown in Fig. 18-23*b*. On the other hand, if the time constant is too large, the output waveform will be too slow to pick up the essential features of the envelope as depicted by the waveform shown in Fig. 18-23*c*. Obviously, a compromise must be

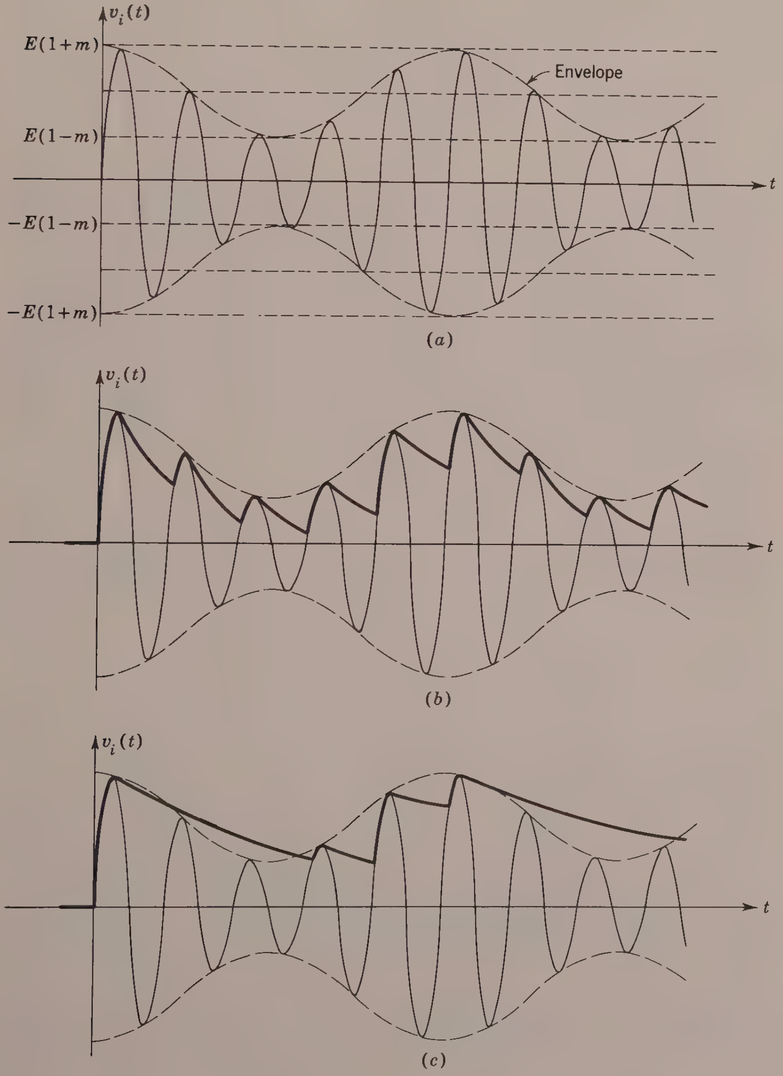


Fig. 18-23. An amplitude-modulated signal and the output waveforms of an envelope detector due to too small or too large a time constant.

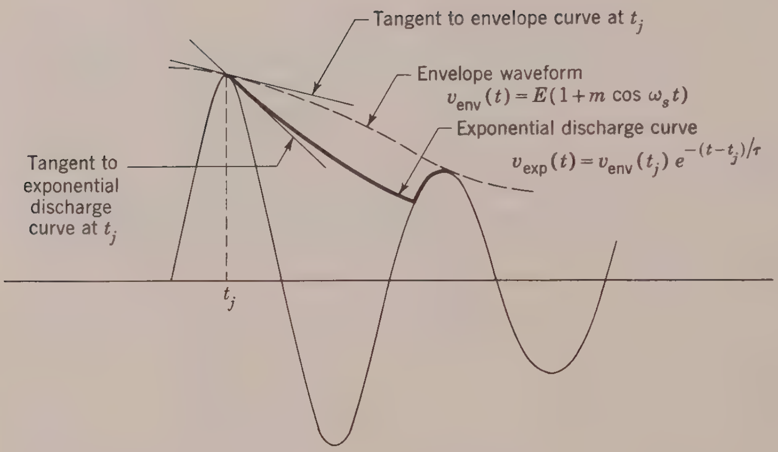

Tangent to envelope curve at t_j
Envelope waveform $v_{\text{env}}(t) = E(1+m\cos\omega_s t)$
Exponential discharge curve $v_{\text{exp}}(t) = v_{\text{env}}(t_j)\,e^{-(t-t_j)/\tau}$
Tangent to exponential discharge curve at t_j
t_j

Fig. 18-24. A blown-up section of the amplitude modulated waveform showing portions of the desired envelope and the exponential discharge curve of the pulse stretcher output waveform.

chosen in order to avoid the ripple distortion and the envelope distortion due to too small and too large a time constant. In order to obtain some design criteria, let us consider a blown-up section of this AM waveform as shown in Fig. 18-24. Clearly, to prevent envelope distortion, the tangent to the exponential discharge curve at the peak point of each pulse of the AM waveform must be *steeper* than the tangent line through the same point of the envelope. If we let t_j be the time of occurrence of each peak, then the slope of the envelope at t_j is given by

$$\left.\frac{dv_{\text{env}}(t)}{dt}\right|_{t=t_j} = \left.\frac{d}{dt}[E(1 + m\cos\omega_s t)]\right|_{t=t_j} = -Em\omega_s \sin\omega_s t_j \tag{18-53}$$

Similarly, the slope of the exponential discharge curve is given by

$$\begin{aligned}\left.\frac{dv_{\text{exp}}(t)}{dt}\right|_{t=t_j} &= \left.\frac{d}{dt}[v_{\text{env}}(t_j)e^{-(t-t_j)/\tau}]\right|_{t=t_j}\\ &= \frac{-v_{\text{env}}(t_j)}{\tau} = -\frac{E(1 + m\cos\omega_s t_j)}{\tau}\end{aligned} \tag{18-54}$$

To avoid envelope distortion, we require

$$\left.\frac{dv_{\text{exp}}(t)}{dt}\right|_{t=t_j} < \left.\frac{dv_{\text{env}}(t)}{dt}\right|_{t=t_j} \tag{18-55}$$

Substituting Eqs. (18-53) and (18-54) into Eq. (18-55), and simplifying, we obtain

$$1 + m[\cos \omega_s t_j - (\omega_s \tau) \sin \omega_s t_j] > 0 \tag{18-56}$$

This equation can be written in the form

$$1 + m\sqrt{1 + (\omega_s \tau)^2}\left[\frac{1}{\sqrt{1 + (\omega_s \tau)^2}} \cos \omega_s t_j - \frac{\omega_s \tau}{\sqrt{1 + (\omega_s \tau)^2}} \sin \omega_s t_j\right] > 0 \tag{18-57}$$

If we define $\tan \theta = \omega_s \tau / 1$, then Eq. (18-57) can be written as

$$1 + m\sqrt{1 + (\omega_s \tau)^2}[\cos (\omega_s t_j + \theta)] > 0 \tag{18-58}$$

Equation (18-58) can be satisfied for all t_j if, and only if,

$$m\sqrt{1 + (\omega_s \tau)^2} < 1 \tag{18-59}$$

Solving for τ from Eq. (18-59), we obtain the desired upper bound

$$\tau < \frac{\sqrt{1 - m^2}}{m\omega_s} \tag{18-60}$$

This relationship can be used as a design criterion for choosing the appropriate value of the time constant $\tau = RC$. Among other things, Eq. (18-60) shows that it is impossible to avoid envelope distortion if the index of modulation m is greater than unity.

18-5-3 THE PULSE-SHARPENING PROPERTY

Instead of smoothing a pulse, there are many applications which call for the opposite operation, namely, sharpening a pulse. We have already indicated in Sec. 18-2 that, whereas an integrator circuit has the pulse-smoothing property, a differentiator circuit has the pulse-sharpening property. In order to obtain better insight into the sharpening mechanism, consider the voltage pulse shown in Fig. 18-25*a* and its time derivative in Fig. 18-25*b*. Observe that the heights of the positive and negative pulses in Fig. 18-25*b* are given by $E/\Delta T$ and $-E/\Delta T$, respectively. Hence, as $\Delta T \to 0$, the pulse height tends to $\pm\infty$ as depicted by the symbols shown in Fig.

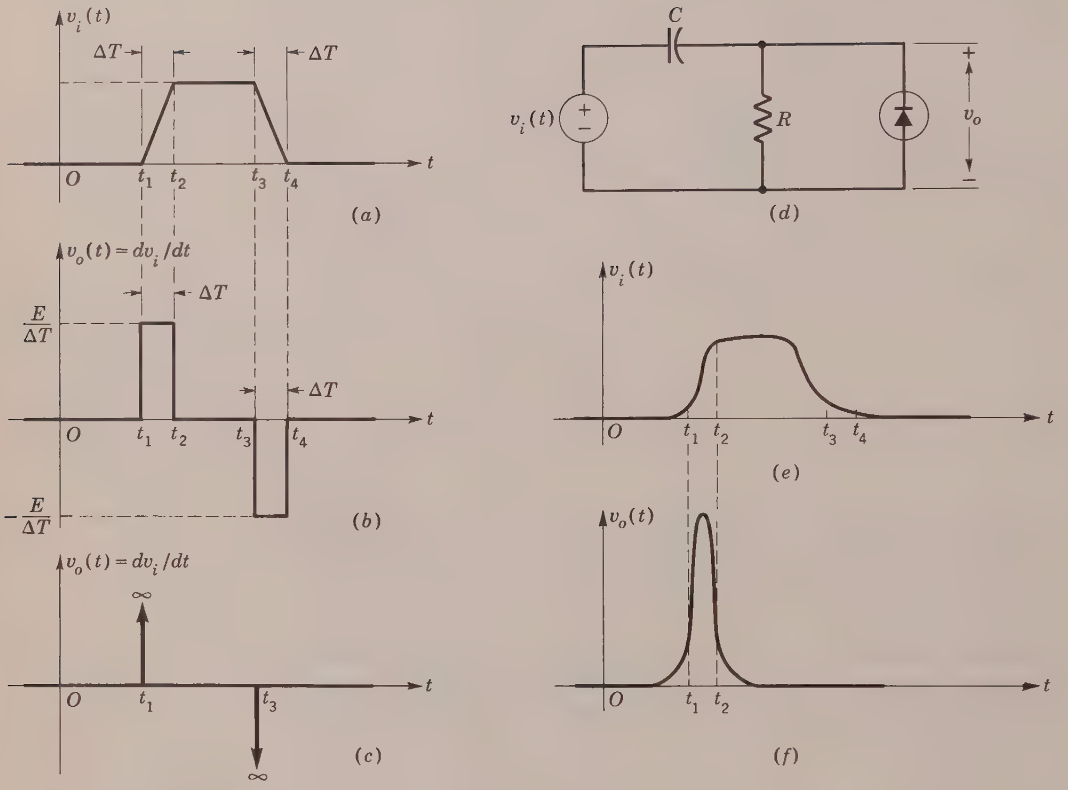


Fig. 18-25. Demonstration of the sharpening property of a pulse-sharpening circuit.

18-25*c*.[1] In practice, ΔT is never zero, and hence the height of the differentiated pulse can be made very large but never infinite.

The above pulse-sharpening property is often used in practice for obtaining very sharp and narrow pulses. This is achieved by clipping off the negative pulse shown in Fig. 18-25*b*. A simple circuit for accomplishing this task consists of connecting a junction diode across an RC differentiator as shown in Fig. 18-25*d*.[2] The input signal $v_i(t)$ is usually chosen to be a short, smooth pulse as shown in Fig. 18-25*e*. The approximate output voltage $v_o(t)$ corresponding to an ideal diode can be easily obtained by differentiating $v_i(t)$ and is shown in Fig. 18-25*f*. The actual waveform will differ slightly because the differentiator is not exact ($\tau \approx 0$), and the diode is not ideal. If necessary, the exact waveform can always be obtained by the constant-slope network approach.

Exercise 1: State the monotonic charging property for the case where the diode terminals are transposed.

Exercise 2: Add an appropriate battery to the clamping circuits shown in

---

[1] Observe that while the pulse height tends to $\pm\infty$, the area under each pulse is equal to E, independent of ΔT. Such a pulse is called an *impulse*.

[2] In this case the time constant RC should be made as small as possible.

Fig. 18-18a and b so that the output waveforms will be clamped at $v_o = 2$ volts and $v_o = -6$ volts, respectively. Draw all pertinent waveforms corresponding to the 10-volt sinusoid shown in Fig. 18-18.

Exercise 3: Repeat Exercise 2 but with the clamping level set at $v_o = -2$ volts and $v_o = 4$ volts, respectively.

Exercise 4: With $C = 10\ \mu$F and with the ideal diode replaced by a type 1NXX2 junction diode, find the exact output voltage $v_o(t)$ for the clamping networks shown in Fig. 18-18. Compare the discrepancy with the ideal case.

Exercise 5: The ripple of a filtered rectifier can be reduced by applying the pulse-stretching property to the full-wave rectifier circuit shown in Fig. 7-8a. Sketch the output waveform of a filtered full-wave rectifier and show that the ripple frequency is equal to twice the frequency of the input sinusoid.

Exercise 6: Show that the upper bound for the time constant of an AM envelope detector corresponding to the information-bearing signal $v_s(t) = E \sin \omega_s t$ is also given by Eq. 18-60.

18-6 SUMMARY

The complete response (solution waveform) of any linear nonautonomous network is always equal to the sum of a zero-input response and a zero-state response. The zero-input response is the solution due only to the initial condition on the energy-storage element (with all sources set equal to zero). The zero-state response is the solution due only to the independent sources (with zero initial condition). This additivity property is called the principle of superposition, and is valid only if the network is linear.

A piecewise-linear network can be analyzed by deriving first the switching plane or the switching space by the iterative piecewise-linear method. Once the switching plane or the switching space is found, the network can be analyzed as a linear nonautonomous network.

Constant-slope network approach This is the most general method available for finding the complete response of any nonautonomous, first-order, nonlinear network. The method consists of obtaining the constant-slope curves by solving for a series of voltage or current waveforms of a resistive constant-slope network.

Practical nonautonomous, first-order networks are designed primarily to transform an input waveform into an output waveform with some desirable properties. They are based mostly on three properties:

1. The monotonic charging property (examples: clamping network, filtered rectifier, voltage doubler)

2. The pulse-smoothing property (examples: pulse stretcher, AM envelope detector)
3. The pulse-sharpening property (example: sharp, narrow-pulse generator)

PROBLEMS

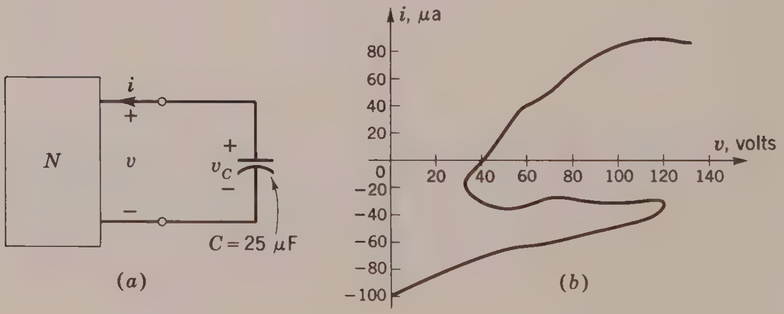
18-1 The constant-slope network approach can be generalized to obtain the solution of autonomous, first-order networks having a multi-valued normal-form equation

$$\frac{dx}{dt} = f(x)$$

where $f(x)$ is a *multivalued* function of x. In this case, certain points on the t-x plane will have more than one director. The pertinent director that applies at a particular point can be decided easily by the inertia postulate.

(a) Sketch the vector field associated with the network shown in Fig. P18-1.

(b) Sketch the solution waveform $v_C(t)$ corresponding to the initial state $v_C(0) = 0$ volt.



(a) (b)

Fig. P18-1.

18-2 Repeat Prob. 18-1 for the network shown in Fig. P18-2.

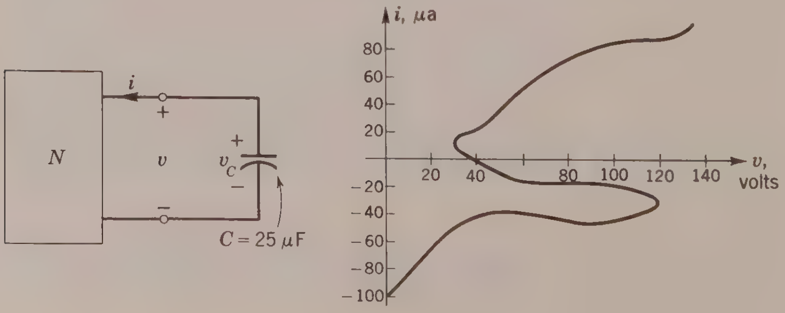


Fig. P18-2.

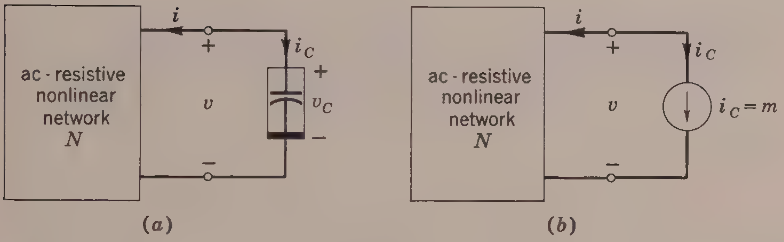
18-3 Consider the nonautonomous, first-order, nonlinear network shown in Fig. P18-3*a*. If the capacitor is charge-controlled, then the normal-form equation is given by

$$\frac{dq}{dt} = f(q,t)$$

where the state variable q denotes the capacitor charge. Show that the constant-slope curve in this case can be found in two steps:

(a) Solve for the $v_C(t,m)$ of the normalized constant-slope network shown in Fig. P18-3*b*.

(b) Perform a graphical composition between $v_C(t,m)$ and the v_C-q curve of the nonlinear capacitor.


ac - resistive nonlinear network N
i
$+$
v
$-$
i_C
$+$
v_C
$-$
ac - resistive nonlinear network N
i
$+$
v
$-$
i_C
$i_C = m$

(*a*) (*b*)

Fig. P18-3.

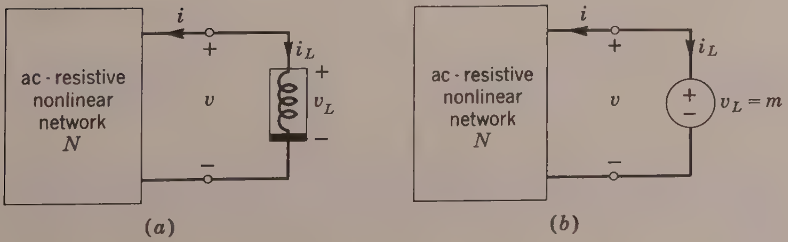
18-4 Consider the nonautonomous, first-order, nonlinear network shown in Fig. P18-4*a*. If the inductor is flux-controlled, then the normal-form equation is given by

$$\frac{d\varphi_L}{dt} = f(\varphi_L,t)$$

where the state variable φ_L denotes the inductor flux linkage. Show that the constant-slope curve in this case can be found in two steps:

(a) Solve for the current $i_L(t,m)$ of the normalized constant-slope network shown in Fig. P18-4*b*.

(b) Perform a graphical composition between $i_L(t,m)$ and the i_L-φ_L curve of the nonlinear inductor.


ac - resistive nonlinear network N
i
$+$
v
$-$
i_L
$+$
v_L
$-$
ac - resistive nonlinear network N
i
$+$
v
$-$
i_L
$v_L = m$

(*a*) (*b*)

Fig. P18-4.

18-5 Consider the clamping circuit shown in Fig. P18-5*a*. If the junction diode is characterized by the *v-i* curve shown in Fig. P18-5*b*, find the output voltage $v_o(t)$ corresponding to a 1-kHz, 150-volt peak-to-peak sinusoid.

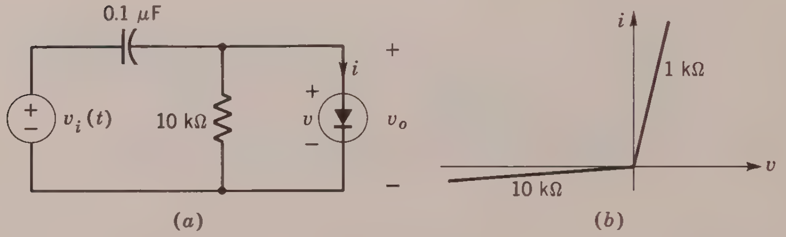


Fig. P18-5.

18-6 The capacitor-diode circuit shown in Fig. P18-6 is often used in practice to convert a low ac input voltage $v_i(t)$ into a high dc output voltage $v_o(t)$. Assuming that $v_i(t)$ is a symmetrical square wave with amplitude equal to E volts, find the *steady-state* value of the voltages $v_1, v_2, \ldots, v_{10}$ and the output voltage v_o.

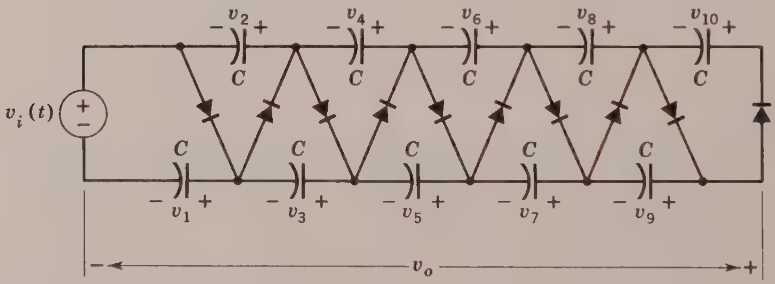


Fig. P18-6.

19 ANALYSIS OF AUTONOMOUS SECOND-ORDER NONLINEAR NETWORKS

19-1 THE PHASE-PLANE TECHNIQUE

The normal-form equations of any autonomous second-order nonlinear network are always given by[1]

$$\frac{dx_1}{dt} = f_1(x_1,x_2) \tag{19-1}$$

$$\frac{dx_2}{dt} = f_2(x_1,x_2) \tag{19-2}$$

where the state variables x_1 and x_2 may represent either voltage or charge for a capacitor, and current or flux linkage for an inductor. It was shown in Sec. 13-5 that any solution of the above equations can be represented by a *trajectory* in the phase plane. Moreover, if the time scale along a trajectory can be found, then the corresponding solution $x_1(t)$ and $x_2(t)$ can be easily plotted as a function of time. Our objectives in this section are, therefore, to develop a systematic technique for finding the trajectory corresponding to any initial state, and then to find the associated time scale.

19-1-1 ISOCLINES AND PHASE PORTRAITS

Suppose we divide Eq. (19-2) by Eq. (19-1) to obtain

$$\frac{dx_2}{dx_1} = \frac{f_2(x_1,x_2)}{f_1(x_1,x_2)} \tag{19-3}$$

This equation defines a unique slope at each point $P(x_1,x_2)$ in the phase plane which a trajectory passing through P must assume.

---

[1] Review the normal-form equations in Sec. 13-4-2.

Hence, if we calculate the slope at each and every point on the phase plane and draw a short director with the calculated slope at the corresponding point, we would obtain the *vector field* associated with Eq. (19-3).† The concept here is exactly analogous to that developed earlier in Sec. 18-4. The only difference is that the time variable t is replaced by the state variable x_1 in Eq. (19-3). Hence, once the vector field is obtained, the trajectory corresponding to any initial state $x_1(t_0)$, $x_2(t_0)$ can be easily sketched.

Just as in Sec. 18-4 there are two general methods for deriving the vector field associated with Eq. (19-3). The first method is the *brute-force method* wherein one systematically substitutes the coordinates of each uniformly spaced point on the phase plane into Eq. (19-3) and calculates the corresponding slope. This method can be implemented automatically with a computer.

A more efficient method consists of finding the constant-slope curves defined by

$$\frac{f_2(x_1,x_2)}{f_1(x_1,x_2)} = m \tag{19-4}$$

where m is a number representing the slope dx_2/dx_1. In order to differentiate between the constant-slope curves of a first-order network and those of a second-order network, the constant-slope curve representing Eq. (19-4) will henceforth be called an *isocline.*

EXAMPLE

Consider the normal-form equations

$$\frac{dx_1}{dt} = x_2 \tag{19-5}$$

$$\frac{dx_2}{dt} = \epsilon(1 - x_1^2)x_2 - x_1 \tag{19-6}$$

Depending on the value of the parameter ϵ, these equations constitute a good mathematical model for many electronic oscillators and multivibrators, and are generally known as the *Van der Pol equation.* The isocline equation is obtained by dividing Eq. (19-6) by Eq. (19-5) and equating the quotient to m; that is,

$$\frac{dx_2}{dx_1} = \frac{\epsilon(1 - x_1^2)x_2 - x_1}{x_2} = m \tag{19-7}$$

Equation (19-7) can be rearranged into the following more explicit form:

---

† Strictly speaking, the elements of this vector field are vectors, and hence each director must be assigned an appropriate direction. The direction of a director at a given point $P(x_1,x_2)$ is determined by the *sign* of dx_1/dt and dx_2/dt at the point $P(x_1,x_2)$. This is found by substituting the value of x_1 and x_2 at P into Eqs. (19-1) and (19-2). For example, if $dx_1/dt > 0$ and $dx_2/dt > 0$ at $P(x_1,x_2)$, then the director at P must be pointing toward the increasing direction of x_1 and x_2; that is, the horizontal and vertical components of this vector must point toward the positive direction of the x_1 axis and the x_2 axis, respectively.

It is important to recognize that Eq. (19-3) determines only the slope of the directors in a vector field. Hence, the system of directors constructed from Eq. (19-3) forms a *linear field,* not a vector field. Observe that if we multiply $f_1(x_1,x_2)$ and $f_2(x_1,x_2)$ by any function $g(x_1,x_2)$ which does not change its sign, then upon dividing Eq. (19-2) by Eq. (19-1), we obtain the same Eq. (19-3). In other words, two distinct autonomous systems of equations may lead to the same linear field, even though each system has a unique vector field.

In spite of the subtle difference between a linear field and a vector field, the term vector field is usually used loosely to represent both cases. Since the context will usually suggest which term is meant, this somewhat ambiguous terminology will be adopted throughout this chapter.

$$x_2 = \frac{x_1}{\epsilon(1 - x_1^2) - m} \tag{19-8}$$

Let us now construct the vector fields corresponding to several typical values of ϵ.

Case 1: $\epsilon = 0$ In this case the network is linear and the isocline equation is given by

$$x_2 = -\frac{1}{m}x_1 \tag{19-9}$$

Hence, the isoclines in this case consist of straight lines through the origin. By drawing a sufficient number of directors with slope m along the corresponding straight-line isocline, we obtain the vector field shown in Fig. 19-1*a*. Observe that for this particular case, all directors are perpendicular to the associated isocline. This means that regardless of where the initial state $[x_1(t_0),x_2(t_0)]$ is, the corresponding trajectory will always be a circle. Hence, the solution waveforms $x_1(t)$ and $x_2(t)$ corresponding to any initial state are always periodic; i.e., there will be no transients.

A typical trajectory corresponding to the initial state $x_1(t_0) = \sqrt{2}$ and $x_2(t_0) = \sqrt{2}$ is shown in Fig. 19-1*a*. In order to determine which direction the trajectory moves as time increases, we must refer back to Eqs. (19-5) and (19-6). At the initial point $P_0(\sqrt{2},\sqrt{2})$ (any other point will do) we found $dx_1/dt > 0$ and $dx_2/dt < 0$. This means that the value of x_1 must increase while that of x_2 must decrease, and hence the trajectory must be in the clockwise direction as shown in Fig. 19-1*a*. It is a simple matter, therefore, to determine the direction of a trajectory. We only have to find the "sign" of dx_1/dt and dx_2/dt at any convenient point P on the trajectory by substituting the coordinates of P into the normal-form equation.

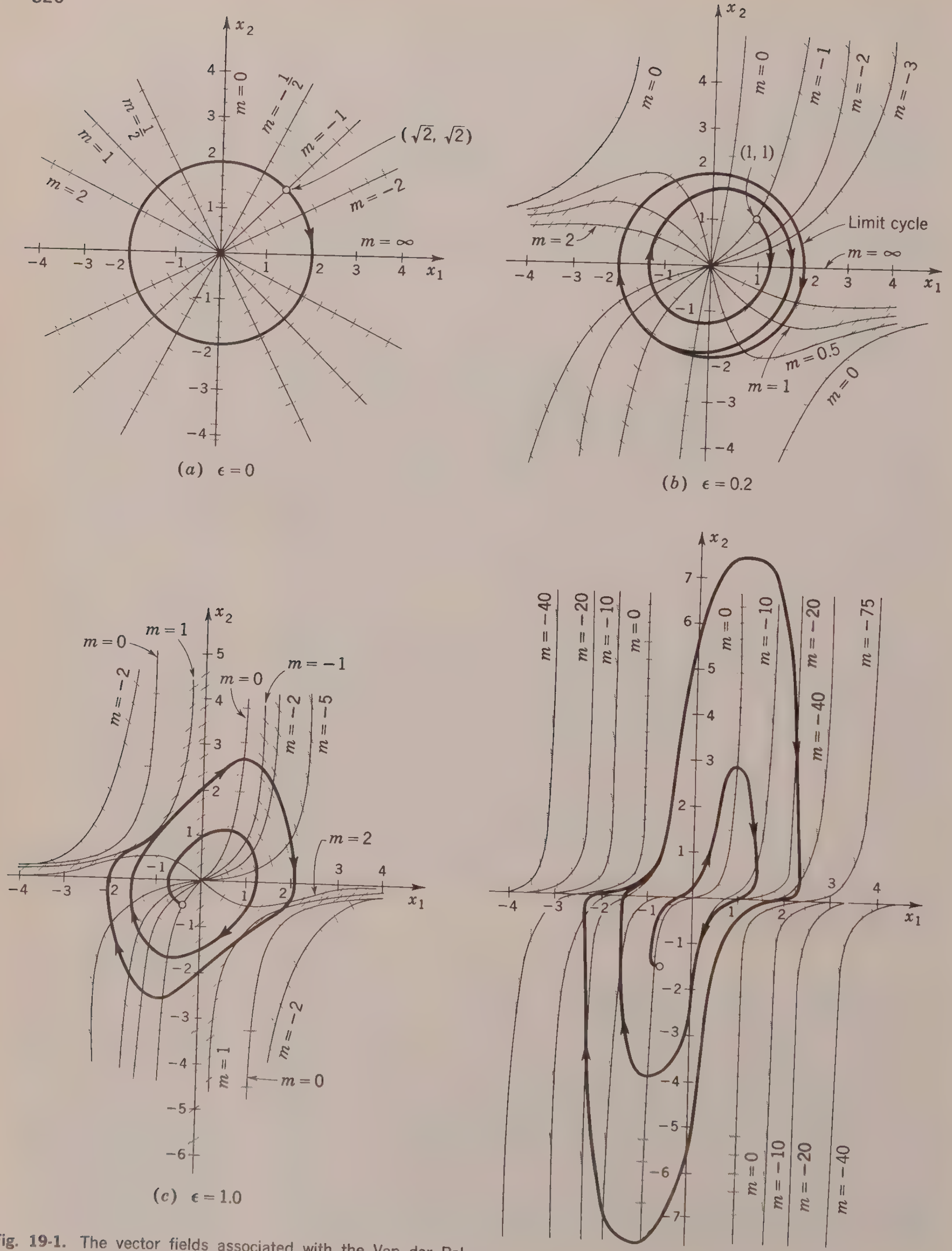
Case 2: $\epsilon = 0.2$ In this case the isocline equation is obtained by substituting $\epsilon = 0.2$ in Eq. (19-8).

$$x_2 = \frac{x_1}{0.2(1 - x_1^2) - m} \tag{19-10}$$

Corresponding to each value m, Eq. (19-10) represents a curve in the phase plane.[1] By plotting a few isoclines and their associated directors, we obtain the vector field shown in Fig. 19-1*b*. A typical trajectory corresponding to the initial state (1,1) is shown on this figure. Observe that this trajectory eventually approaches a closed curve. This means that the solution waveforms $x_1(t)$ and $x_2(t)$ tend

---

[1] The curve representing an isocline may consist of several separate branches. For the case $\epsilon = 0.2$, each isocline corresponding to $m < 0.2$ will contain three branches.



(a) $\epsilon = 0$

(b) $\epsilon = 0.2$

(c) $\epsilon = 1.0$

(d) $\epsilon = 5.0$

Fig. 19-1. The vector fields associated with the Van der Pol equation corresponding to several typical values of ϵ.

to a periodic waveform in the steady state. A careful inspection of this vector field shows that regardless of where the initial state is, all trajectories must eventually converge to the same closed curve. Such a limiting curve is called a *limit cycle.*[1] It can be shown that only nonlinear systems can have limit cycles.

Case 3: $\epsilon = 1.0$ In this case the isocline equation is given by

$$x_2 = \frac{x_1}{(1 - x_1^2) - m} \tag{19-11}$$

The vector field associated with this equation is shown in Fig. 19-1*c*, together with a typical trajectory. Observe that this trajectory, as well as any other, eventually converges toward a limit cycle.

Case 4: $\epsilon = 5.0$ In this case the isocline equation is given by

$$x_2 = \frac{x_1}{5(1 - x_1^2) - m} \tag{19-12}$$

The vector field associated with this equation is shown in Fig. 19-1*d*, together with a typical trajectory. Once again it is easy to see that all trajectories converge to a limit cycle.

A comparison of the vector fields corresponding to different values of ϵ shows that the limit cycle is almost circular when ϵ is small; however, as ϵ increases, parts of the limit cycle become steeper, it will eventually approach a vertical line when $\epsilon = \infty$.†

Although the isoclines are different for different networks, there are some general properties which are common to all of them. For example, an examination of Eq. (19-3) shows that if both $f_1(x_1,x_2)$ and $f_2(x_1,x_2)$ are single-valued functions, then each point $P(x_1,x_2)$ for which $f_1(x_1,x_2) \neq 0$ and $f_2(x_1,x_2) \neq 0$ is associated with a *unique slope.* Such points will, henceforth, be called *regular points.* What about points $P(\bar{x}_1,\bar{x}_2)$ for which both $f_1(\bar{x}_1,\bar{x}_2) = f_2(\bar{x}_1,\bar{x}_2) = 0$? An examination of Eqs. (19-1) and (19-2) shows that these points must correspond to the equilibrium states, also known as *singular points,* defined in Sec. 13-6. Observe that $dx_2/dx_1 = 0/0$ at an equilibrium state, and is therefore *indeterminate.* This means that unlike the regular points, an equilibrium state may possess any slope. Summarizing the above observations, we conclude that *the isoclines of any autonomous nonlinear network cannot intersect each other except at an equilibrium state.* This conclusion provides us with a simple means for locating all equilibrium states of a nonlinear network. For example, an inspection of the vector fields corresponding to the Van der Pol equation immediately shows that the origin is the only equilibrium state. Moreover, observe that this equilibrium state is *unstable* for all $\epsilon > 0$.

---

[1] Not all closed trajectories are limit cycles. To qualify as a limit cycle, all trajectories in the vicinity of the closed trajectory must not be closed. Hence, the circular trajectories in Fig. 19-1*a* are *not* limit cycles.

† In certain classes of second-order electronic oscillators which are described by the Van der Pol equation, the case $\epsilon = \infty$ corresponds to the limiting case in which the oscillator degenerates into a first-order network upon neglecting a parasitic inductor or a parasitic capacitor. The vertical trajectory in this case corresponds to the jump phenomenon described in Sec. 14-6.

19-1-2 TIME SCALING THE TRAJECTORIES

In order to recover the waveforms $x_1(t)$ and $x_2(t)$ associated with a trajectory, we must first provide the time scale along the trajectory. Consider the typical trajectory Γ shown in Fig. 19-2*a*. Given the initial time t_0, our objective will be to calculate $t_1, t_2, \ldots, t_{j-1}, t_j, \ldots$ etc., along a set of conveniently spaced points on Γ.

Consider first how to calculate t_1. Let that portion of the trajectory between t_0 and t_1 be denoted by $x_2 = \Gamma(x_1)$.† Then Eq. (19-1) can be written as

$$dt = \frac{dx_1}{f_1(x_1,\Gamma(x_1))} \tag{19-13}$$

Integrating both sides of this equation (from t_0 to t_1 on the left, and from $x_1(t_0)$ to $x_1(t_1)$ on the right), we obtain

$$t_1 = t_0 + \int_{x_1(t_0)}^{x_1(t_1)} \frac{dx_1}{f_1(x_1,\Gamma(x_1))} \tag{19-14}$$

The integrand on the right of Eq. (19-14) can be represented as a curve

$$y = \frac{1}{f_1(x_1,\Gamma(x_1))} \tag{19-15}$$

in the x_1-y plane from $x_1(t_0)$ to $x_1(t_1)$ as shown in Fig. 19-2*b*. Hence, the area under this curve is equal numerically to the time it takes the trajectory Γ to move from t_0 to t_1. Notice that if this curve is almost linear, then Eq. (19-14) can be approximated by

$$t_1 \approx t_0 + \frac{\Delta x_1}{f_1(\bar{x}_1,\bar{x}_2)} \tag{19-16}$$

where $(\bar{x}_1,\bar{x}_2)$ is some point on Γ between t_0 and t_1. An examination of Fig. 19-2*b* shows that in order for t_1 to approach the exact value, $\bar{x}_1$ must be chosen such that the two shaded areas cancel each other. If Δx_1 is small, a good approximation is given by $\bar{x}_1 = \frac{1}{2}[x_1(t_0) + x_1(t_1)]$ and $\bar{x}_2 = \frac{1}{2}[x_2(t_0) + x_2(t_1)]$. Hence, t_1 can

† If we choose the spacings in such a way that the trajectory between t_{j-1} and t_j approximates a straight line, then $x_2 = k_1x_1 + k_2$, where k_1 and k_2 are the slope and intercept.

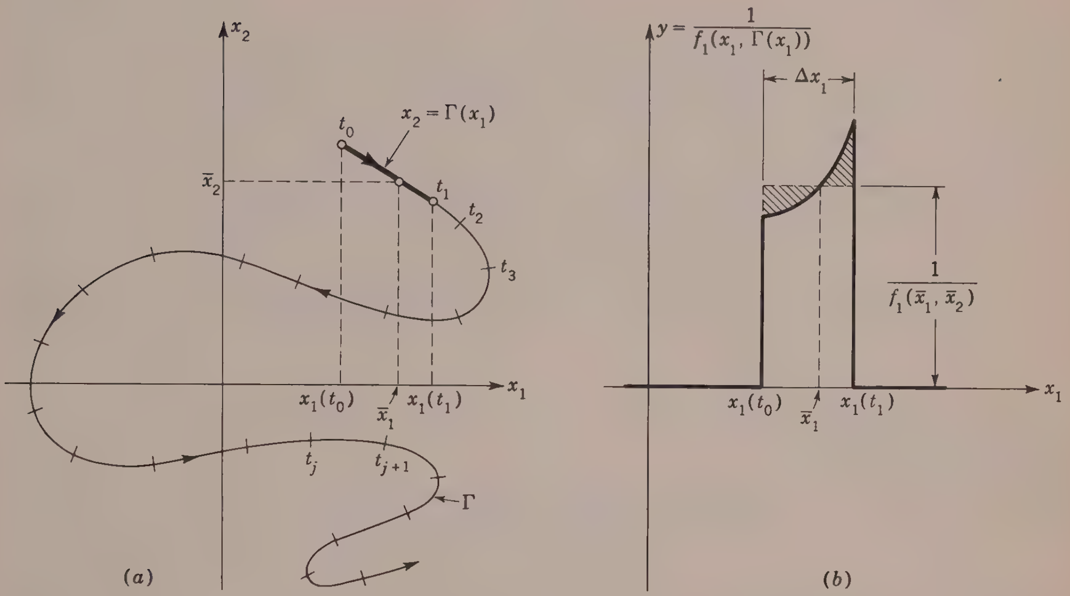


Fig. 19-2. A typical trajectory whose time scale is to be found.

be calculated either exactly by Eq. (19-14) or approximately by Eq. (19-16). Knowing t_1, we can calculate t_2, t_3, . . . etc., by simply changing the subscripts of t in Eqs. (19-14) or (19-16).

An inspection of Eqs. (19-14) and (19-16) shows that since $f_1(x_1,x_2) = 0$ at an equilibrium state Q, it will take an infinite time interval for the trajectory to arrive at Q. Hence, we conclude that *no trajectory may arrive at an equilibrium state in finite time.*

It is possible to obtain more qualitative information if the normal-form equations assume the special form

$$\frac{dx_1}{dt} = x_2 \tag{19-17}$$

$$\frac{dx_2}{dt} = f_2(x_1,x_2) \tag{19-18}$$

and if the time spacings on the trajectory have been chosen close enough so that the trajectory between adjacent points can be approximated by a straight-line segment as shown in Fig. 19-3*a*; namely,

$$x_2 = x_2(t_j) + m[x_1 - x_1(t_j)] \tag{19-19}$$

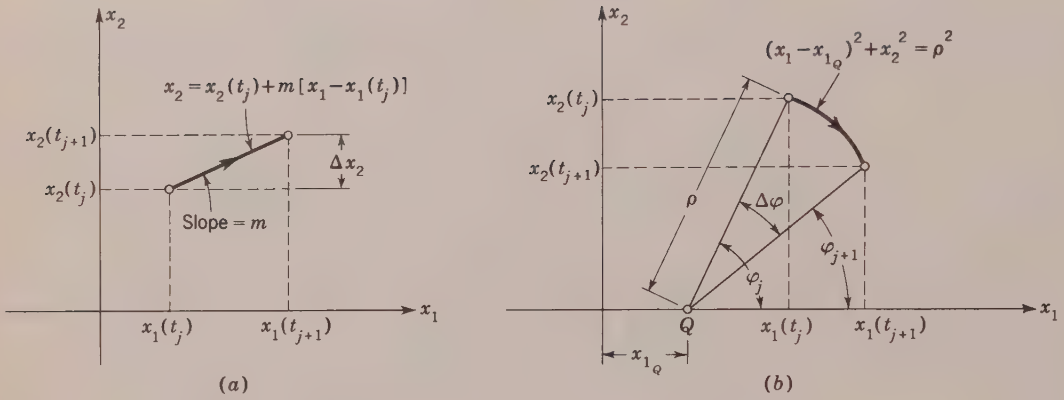


Fig. 19-3. Two simple trajectory segments consist of either a straight-line segment or a circular arc.

where m is the slope of the straight-line segment. Using Eqs. (19-14), (19-17), and (19-19), we obtain

$$t_{j+1} = t_j + \int_{x_1(t_j)}^{x_1(t_{j+1})} \frac{dx_1}{x_2(t_j) + m[x_1 - x_1(t_j)]} \tag{19-20}$$

Integrating Eq. (19-20) and simplifying, we obtain

$$t_{j+1} = t_j + \frac{1}{m} \ln \frac{x_2(t_{j+1})}{x_2(t_j)} \tag{19-21}$$

Equation (19-21) is very useful because with it we can calculate the time parameter at any point on a straight-line trajectory of any network described by Eqs. (19-17) and (19-18). An examination of Eq. (19-21) reveals that the time interval $\Delta t \equiv t_{j+1} - t_j$ decreases as m increases and approaches zero as $m \to \infty$ (assuming that the trajectory traverses the same vertical distance). In other words, the steeper the trajectory, the less time it takes to move from one point on the trajectory to another. Hence, it takes zero time to move along a vertical trajectory.

What happens if $m = 0$? In this case Eq. (19-21) becomes indeterminate because when $m = 0$, the trajectory is horizontal, and hence $x_2(t_{j+1}) = x_2(t_j)$, and $\ln [x_2(t_{j+1})/x_2(t_j)] = 0$. However, it is easy to obtain the correct expression directly from Eq. (19-20) with $m = 0$; namely,

$$t_{j+1} = t_j + \frac{x_1(t_{j+1}) - x_1(t_j)}{x_2(t_j)} \tag{19-22}$$

This equation is, of course, valid only for a horizontal trajectory of Eqs. (19-17) and (19-18). Observe that Eq. (19-22) shows that it takes less time to traverse a horizontal trajectory farther from the x_1 axis than a horizontal trajectory of equal length but located nearer the x_1 axis.

If the curvature of a portion of a trajectory is large, it will be more accurate to approximate it by a *circular arc* as shown in Fig. 19-3*b*. If the center of the arc is located at Q and if the radius is ρ, then under the assumption that the network's normal-form equations are given by Eqs. (19-17) and (19-18), we can rewrite Eq. (19-14) as follows:

$$t_{j+1} = t_j + \int_{x_1(t_j)}^{x_1(t_{j+1})} \frac{dx_1}{\sqrt{\rho^2 - (x_1 - x_{1_Q})^2}} \tag{19-23}$$

Integrating and simplifying this equation, we obtain

$$\begin{aligned} t_{j+1} &= t_j + \cos^{-1}\left[\frac{x_1(t_j) - x_{1_Q}}{\rho}\right] - \cos^{-1}\left[\frac{x_1(t_{j+1}) - x_{1_Q}}{\rho}\right] \\ &= t_j + \varphi_j - \varphi_{j+1} \end{aligned}$$

or

$$t_{j+1} - t_j = \Delta\varphi \tag{19-24}$$

This equation reveals a most surprising fact that the time it takes to traverse a circular arc trajectory (with center on the x_1 axis) of Eqs. (19-17) and (19-18) does not depend on the length of the arc, but rather on the angle $\Delta\varphi$ (in radians) subtended by the arc.

As an application of these properties, let us consider the Van der Pol equation again (Eqs. 19-5 and 19-6). Notice that this equation has the same form as Eqs. (19-17) and (19-18). Corresponding to the circular trajectory in Fig. 19-1*a*, the solution waveform, we immediately conclude, is a sinusoid with period $T = 2\pi$ (since $\Delta\varphi = 2\pi$ for a complete circle) as shown in Fig. 19-4*a*. With the help of Eq. (19-16) we obtain the approximate solution waveforms $x_2(t)$ shown in Fig. 19-4*b* to *d* corresponding to the trajectories shown in Fig. 19-1*b* to *d*. As expected, we find that the steeper the trajectory, the faster is the waveform for $x_2(t)$. In the limit where $\epsilon \to \infty$, the waveform $x_2(t)$ approaches a square wave.

It is clear that even with the help of Eqs. (19-21) and (19-24), the time-scaling procedure is at best an approximate and tedious thing to do. Actually, we seldom do this in practice, for if we were

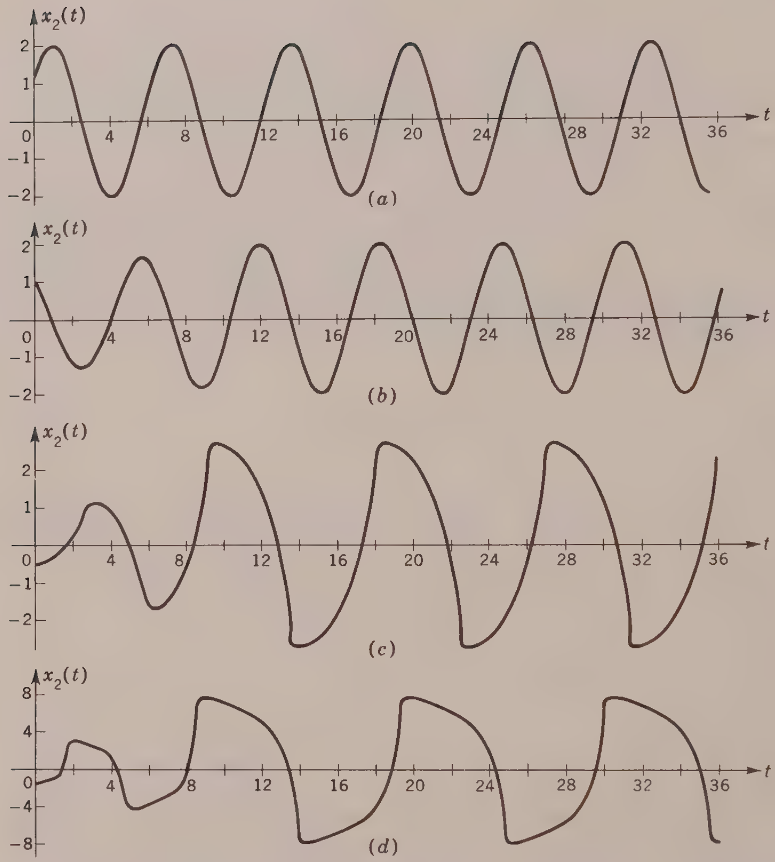


Fig. 19-4. The solution waveforms corresponding to the trajectories in Fig. 19-1.

interested only in the solution waveform $x_2(t)$, there are other more efficient methods, such as numerical techniques. The main value of the above results is the qualitative properties which we learned concerning the behavior of trajectories. For example, we learned that it takes an infinite time to arrive at an equilibrium state. We observe that for networks described by Eqs. (19-17) and (19-18), it takes less time to traverse a steep trajectory, and in the limit, it takes zero time to traverse a vertical trajectory. We also saw that it takes the same time for two trajectories of Eqs. (19-17) and (19-18) to traverse two concentric circular arcs (with center on the x_1 axis) which are subtended by the same angle, and if an arc consists of a complete circle, the corresponding solution is a sinusoidal waveform with period $T = 2\pi$.

Exercise 1: Construct the vector field for the Van der Pol equation corresponding to $\epsilon = 0.5$.

Exercise 2: Show that the isoclines of the Van der Pol equation corresponding to $m < \epsilon$ consist of three separate branches.

Exercise 3: Construct the vector field for the Van der Pol equation corresponding to $\epsilon = -1.0$. Locate the equilibrium state and determine whether it is stable or unstable.

Exercise 4: Show that if the trajectory Γ is expressed by $x_1 = \Gamma(x_2)$, then the time scale t_j can be exactly calculated from

$$t_{j+1} = t_j + \int_{x_2(t_j)}^{x_2(t_{j+1})} \frac{dx_2}{f_2(x_2,\Gamma(x_2))}$$

or approximately from

$$t_{j+1} \approx t_j + \frac{\Delta x_2}{f_2(\bar{x}_1,\bar{x}_2)}$$

where $(\bar{x}_1,\bar{x}_2)$ is some point on Γ between t_j and t_{j+1}.

Exercise 5: Derive Eqs. (19-21) and (19-24).

Exercise 6: Show that any second-order differential equation of the form $d^2x/dt^2 = f(x,dx/dt)$ can be transformed into the form of Eqs. (19-17) and (19-18) by defining $x_1 = x$ and $x_2 = dx_1/dt$.

19-2 THE ISOCLINE NETWORK METHOD

The phase-plane technique developed in the preceding section is a powerful and very useful method for obtaining the qualitative behavior of the solutions associated with Eqs. (19-1) and (19-2). Unfortunately, it is usually difficult, if not impractical, to obtain the normal-form equations of practical networks in *analytic form.*[1] The basic difficulty is that most practical nonlinear elements are characterized graphically. Our objective in this section is to develop a method for drawing the isoclines without deriving the isocline equation [Eq. (19-3)]. The key to this problem lies in the definition of an *isocline network.*

19-2-1 THE ISOCLINE NETWORK

Consider first the second-order capacitor network shown in Fig. 19-5*a*.† The equations of motion are given by

$$\frac{dv_{C_1}}{dt} = -\frac{i_1}{C_1} \tag{19-25}$$

$$\frac{dv_{C_2}}{dt} = -\frac{i_2}{C_2} \tag{19-26}$$

---

[1] We assume that all networks considered in this chapter possess a well-defined normal-form equation, even though it may not be possible to write the equations in explicit analytic form.

† For convenience we shall consider only the case where both capacitors are linear. The extension to include nonlinear capacitors will be given as problems.

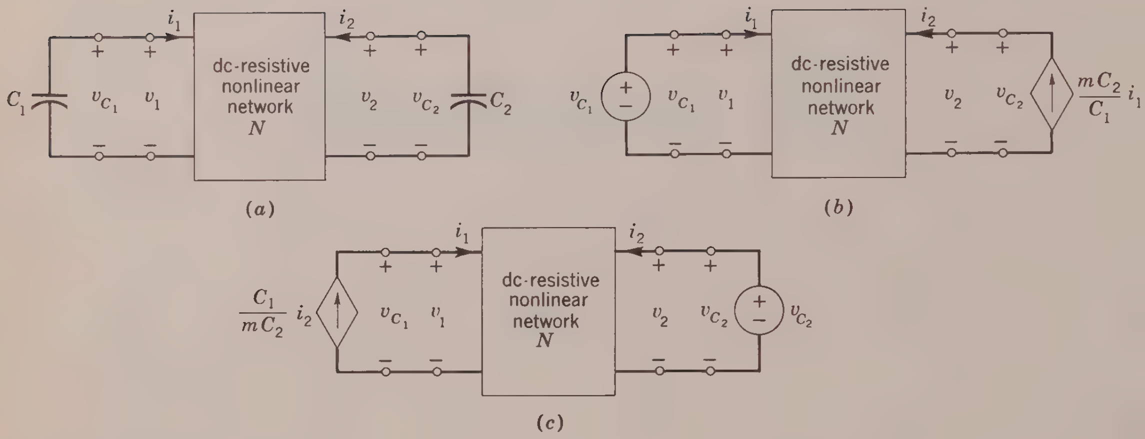


Fig. 19-5. A second-order capacitor network and its associated isocline networks.

Dividing Eq. (19-26) by Eq. (19-25), we obtain

$$\frac{dv_{C_2}}{dv_{C_1}} = \frac{C_1 i_2}{C_2 i_1} \tag{19-27}$$

An isocline corresponding to slope m is, by definition, the locus of all points in the v_{C_1}-v_{C_2} plane for which

$$\frac{dv_{C_2}}{dv_{C_1}} = m \tag{19-28}$$

Equating Eqs. (19-27) and (19-28) and solving either for i_2 or for i_1, we obtain

$$i_2 = \left(\frac{mC_2}{C_1}\right) i_1 \tag{19-29}$$

or

$$i_1 = \left(\frac{C_1}{mC_2}\right) i_2 \tag{19-30}$$

Equations (19-29) and (19-30) can be interpreted as the constraints that must be satisfied by the currents i_1 and i_2 in order for the voltages v_{C_1} and v_{C_2} to fall on the isocline of slope m. The constraint imposed by Eq. (19-29) can be "simulated" by connecting a voltage source across port 1 and a current-controlled current source across port 2 of the resistive subnetwork as shown in Fig. 19-5*b*. Similarly, the constraint imposed by Eq. (19-30) can be "simulated" by the resistive network shown in Fig. 19-5*c*. Observe that the network in Fig. 19-5*b* is being driven at port 1 by an independent voltage source. The solution v_{C_2} corresponding to each value of v_{C_1} must, therefore, be a point on the isocline. But by definition this is simply the v_{C_2}-vs.-v_{C_1} TC plot. Similarly, since we are driving port 2 of the network in Fig. 19-5*c* by a voltage source, the v_{C_1}-vs.-v_{C_2} TC plot must coincide with the isocline of slope m. In view of the above interpretations, we shall call the resistive networks in Fig. 19-5*b* and *c* the *isocline network 1* and the *isocline network 2*, respectively.[1] Let us summarize the above results in the form of a theorem.

CAPACITOR-LOADED ISOCLINE NETWORK THEOREM[2]

The isocline corresponding to slope m of the capacitor-loaded network shown in Fig. 19-5*a* is identical with the v_{C_2}-vs.-v_{C_1} TC plot of the isocline network 1 shown in Fig. 19-5*b* or the v_{C_1}-vs.-v_{C_2} TC plot of the isocline network 2 shown in Fig. 19-5*c*.

The above theorem provides us with a practical method for constructing the vector field associated with the network shown in Fig. 19-5*a*. Observe that since the isocline network is resistive, we can apply the techniques developed in Part 2 of this book for determining the TC plots. Hence, once again we have transformed a difficult dynamic-network problem into a simpler resistive-network problem. Since the TC plots of the isocline network are used only for obtaining the isoclines, they do not represent physically measurable quantities. In fact, these TC plots are generally multivalued and may contain several separate branches. If the dynamic network has a unique solution, i.e., if its normal-form equations are single-valued, then the intersections of the TC plots of the isocline network are the equilibrium states. Another useful property can be found by an inspection of the isocline networks in Fig. 19-5*b* and *c*. Observe that the controlled current source of the isocline network 1 becomes an open circuit when $m = 0$, and that of the isocline network 2 becomes an open circuit when $m = \infty$. Hence, the $m = 0$ and the $m = \infty$ isoclines of the capacitor-loaded network are identical with the v_{C_2}-vs.-v_{C_1} and the v_{C_1}-vs.-v_{C_2} TC plot of the resistive subnetwork N (without the controlled source), respectively.

---

[1] The indexes 1 and 2 refer to the port number which is being driven by the *independent* source.

[2] For a more detailed discussion of related results, see L. O. Chua, Two New Theorems in Nonlinear Networks, *Midwest Symp. Circuit Theory, 11th, University of Notre Dame, South Bend, Ind.*, pp. 340–351, May 13–14, 1968.

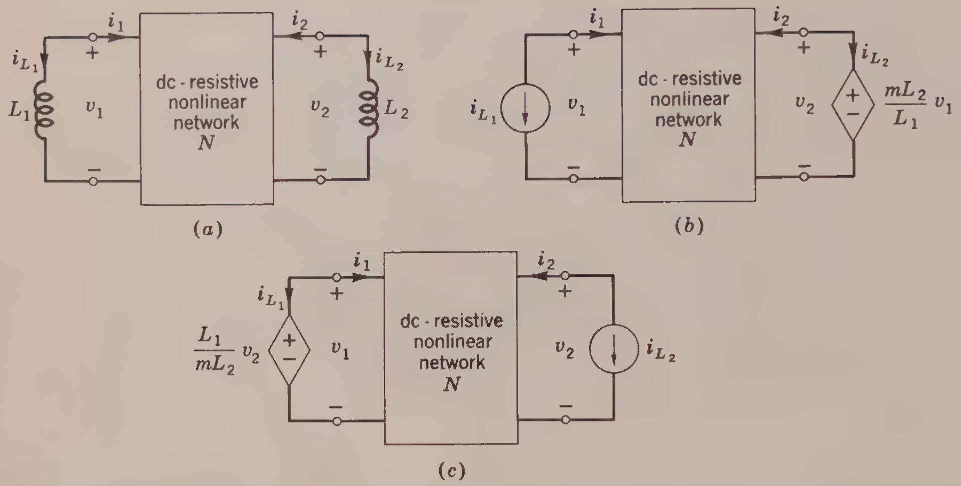


Fig. 19-6. A second-order inductor network and its associated isocline networks.

By a similar procedure, we can state the following analogous theorems:

INDUCTOR-LOADED ISOCLINE NETWORK THEOREM

The isocline corresponding to slope *m* of the inductor-loaded network shown in Fig. 19-6*a* is identical with the i_{L_2}-vs.-i_{L_1} TC plot of the isocline network 1 shown in Fig. 19-6*b* or the i_{L_1}-vs.-i_{L_2} TC plot of the isocline network 2 shown in Fig. 19-6*c*.

CAPACITOR-INDUCTOR-LOADED ISOCLINE NETWORK THEOREM

The isocline corresponding to slope *m* of the capacitor-inductor-loaded network shown in Fig. 19-7*a* is identical with the v_{C_2}-vs.-i_{L_1} TC plot of the isocline network 1 shown in Fig. 19-7*b* or the i_{L_1}-vs.-v_{C_2} TC plot of the isocline network 2 shown in Fig. 19-7*c*.

19-2-2 SOME PRACTICAL NETWORKS

Let us now apply the above theorems to analyze some practical networks.

EXAMPLES

1. Consider the tunnel-diode circuit shown earlier in Fig. 13-18*c*. The normal-form equations of this network were derived in Sec. 13-6-1 and are given by Eqs. (13-86) and (13-87). Dividing

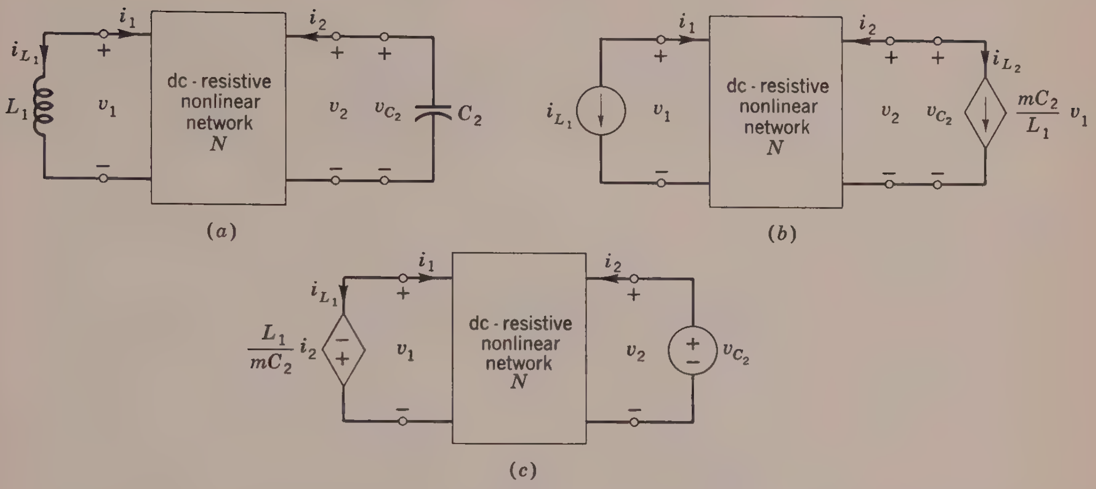


Fig. 19-7. A second-order inductor-capacitor network and its associated isocline networks.

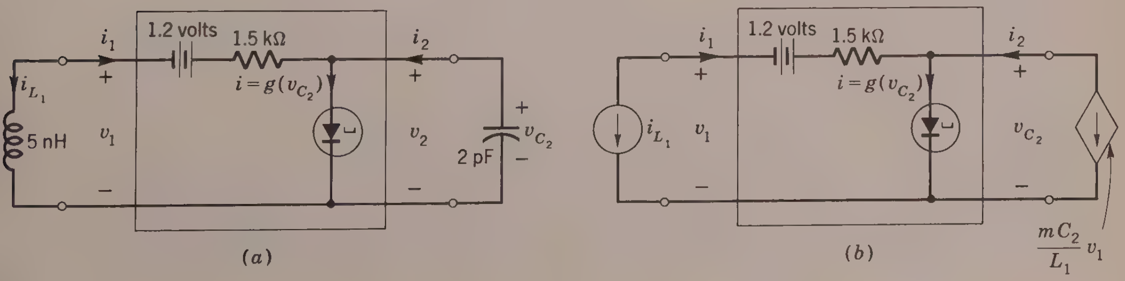
Eq. (13-87) by Eq. (13-86) and simplifying, we obtain the isocline equation

$$\frac{dv_{C_2}}{di_{L_1}} = 2.5 \times 10^3\left[\frac{-i_{L_1} - g(v_{C_2})(10^{-3})}{v_{C_2} - 1.5(10^3)i_{L_1} - 1.2}\right] = m \qquad (19\text{-}31)$$

where $g(v_{C_2})$ is defined by Eq. (13-88). Since Eq. (19-31) is already in analytic form, we could construct the vector field associated with this network directly using this equation. Before we do this, however, let us verify that the isocline network approach would lead to the same answer.

Let us first redraw the network in Fig. 13-18*c* into the form shown in Fig. 19-8*a*. The associated *isocline network 1* is shown in Fig. 19-8*b*. According to the capacitor-inductor-loaded isocline network theorem, the isoclines of this network are identical with the v_{C_2}-vs.-i_{L_1} TC plots. Hence, if $g(v_{C_2})$ had not been given analytically, we could obtain these isoclines by simply solving for the TC plots corresponding to different

Fig. 19-8. A tunnel-diode switching circuit and its associated isocline network 1.



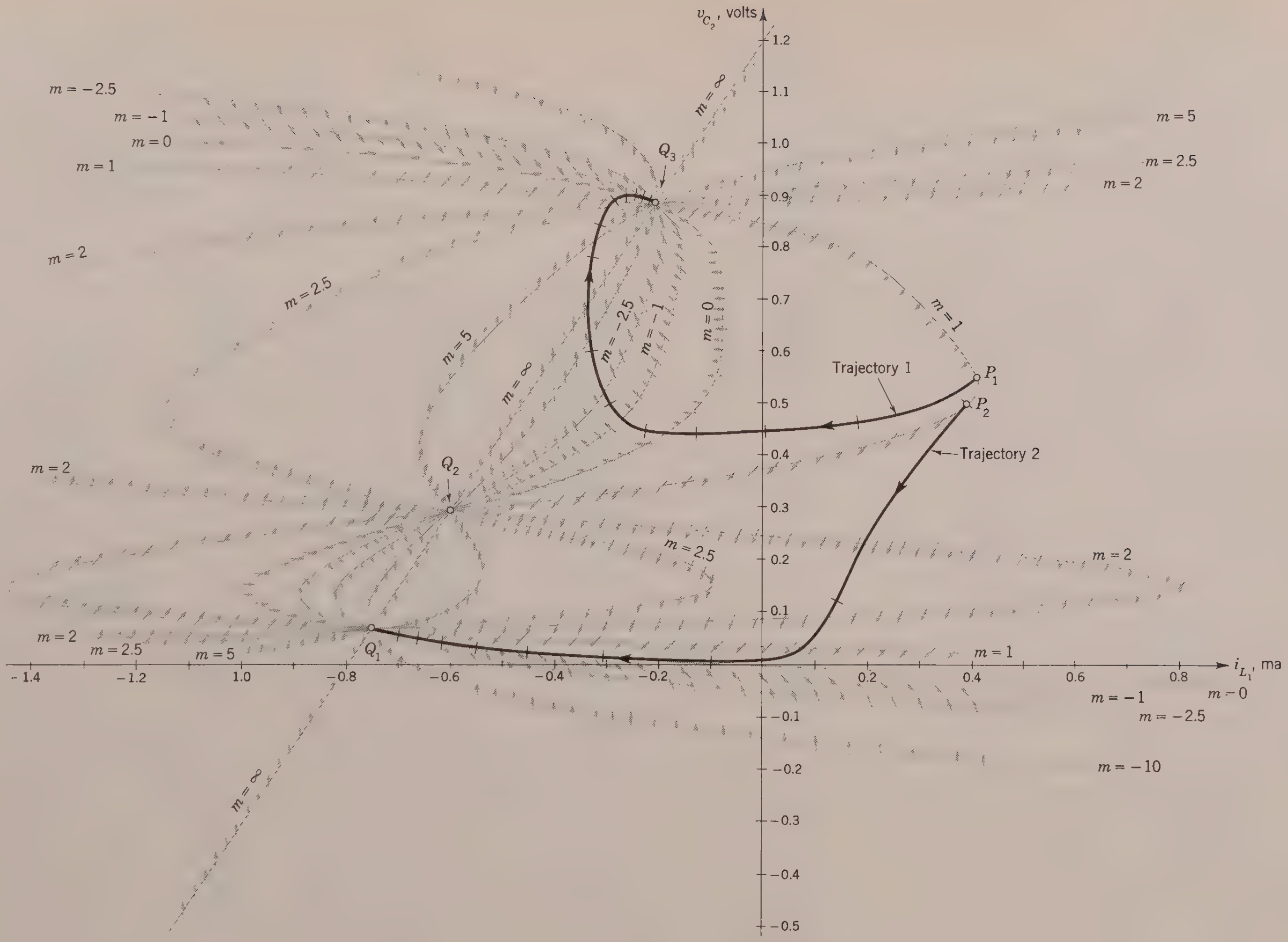


Fig. 19-9. The vector field and trajectories associated with the tunnel-diode switching circuit of Fig. 19-8*a*.

values of m by the methods developed in Part 2 of this book or by the automatic computer program Meca described in Sec. 12-5-2. Since $g(v_{C_2})$ is given analytically, let us derive the TC plots analytically. By writing KVL and KCL equations, we obtain

$$v_1 = v_{C_2} - 1.5(10^3)i_{L_1} - 1.2 \tag{19-32}$$

$$-i_{L_1} - g(v_{C_2})(10^{-3}) = \left(\frac{mC_2}{L_1}\right)v_1 \tag{19-33}$$

Substituting Eq. (19-32) for v_1 in Eq. (19-33) and simplifying, we obtain the equation of the TC plot.

$$\frac{L_1}{C_2}\left[\frac{-i_{L_1} - g(v_{C_2})(10^{-3})}{v_{C_2} - 1.5(10^3)i_{L_1} - 1.2}\right] = m \tag{19-34}$$

where $L_1/C_2 = (5 \times 10^{-9})/(2 \times 10^{-12}) = 2.5 \times 10^3$. Observe that Eqs. (19-34) and (19-31) are identical, as they should.

By assigning different values to m in Eq. (19-31), we obtain the vector field shown in Fig. 19-9. The "sense" of the arrowheads associated with the directors at each point $v_{C_2} = E$, $i_{L_1} = I$ is determined by the "sign" of dv_{C_2}/dt and di_{L_1}/dt. This in turn is determined by the sign of the current i_2 and the voltage v_1 of the resistive network obtained by replacing the capacitor with a voltage source $v_2 = E$ and the inductor with a current source $i_1 = -I$. The systematic method, of course, consists of finding the sense of all directors lying on each isocline. This consists of obtaining either the i_2-vs.-i_{L_1} TC plot or the v_1-vs.-i_{L_1} DP plot of the isocline network shown in Fig. 19-8*b* and then determining the intervals of i_{L_1} for which this plot lies above the i_{L_1} axis.[1] Observe that all isoclines intersect at the three equilibrium states Q_1, Q_2, and Q_3 obtained earlier in Fig. 13-19. An examination of the vector field shown in Fig. 19-9 shows that Q_1 and Q_3 are stable, but Q_2 is unstable. The two trajectories shown earlier in Fig. 13-19 can now be trivially obtained as shown in Fig. 19-9. For the sake of illustration, these two trajectories have been time-scaled by the methods described earlier. The complete waveforms of $i_{L_1}(t)$ and $v_{C_2}(t)$ corresponding to these two trajectories are shown in Fig. 19-10.

2. Consider next the twin tunnel-diode switching circuit shown in Fig. 19-11*a*. The two tunnel diodes are assumed to be characterized by the same piecewise-linear v_d-i_d curve shown in Fig. 19-11*b*. The associated isocline network 1 is shown in Fig. 19-11*c*. Since $C_1 = C_2 = 10$ $\mu\mu$F, the coefficient of the current-controlled current source is equal to m. The TC plots of this

[1] Since the slope of the directors associated with each isocline is known a priori, *the time rate of change of only one of the two state variables is sufficient to determine the "sense" of the directors.* Hence, either the i_2-vs.-i_{L_1} TC plot or the v_1-vs.-i_{L_1} DP plot is sufficient to determine the sense of the directors.

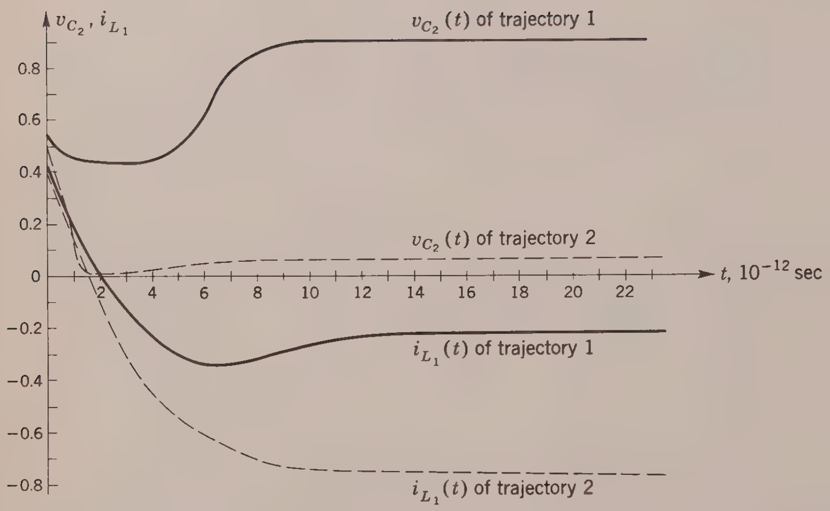
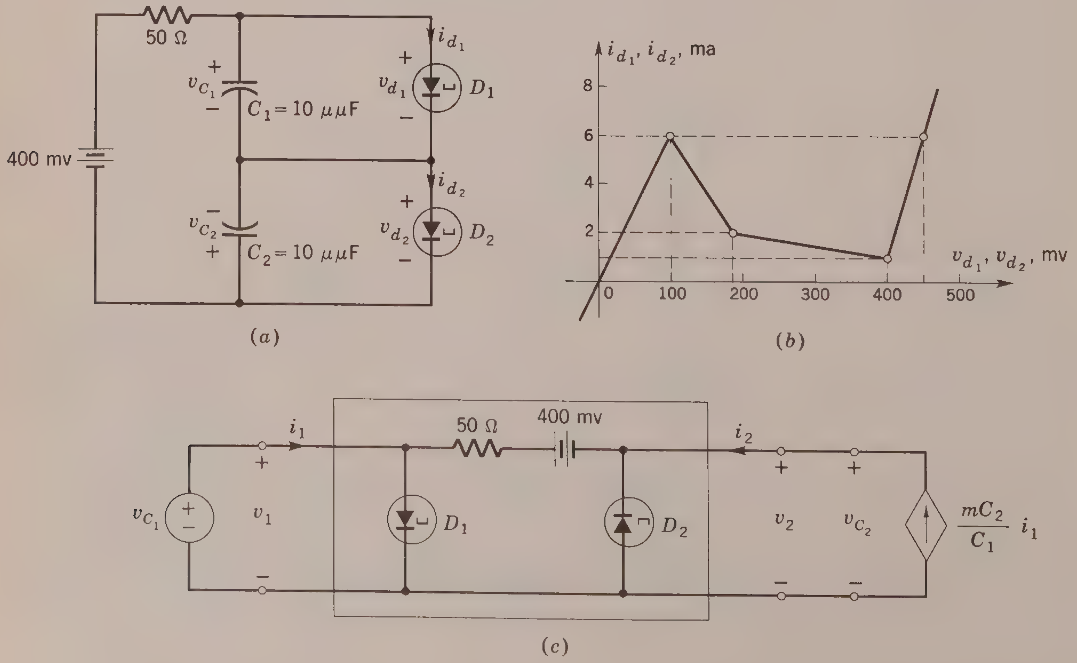


Fig. 19-10. The waveforms of $i_{L_1}(t)$ and $v_{C_2}(t)$ corresponding to the two trajectories shown in Fig. 19-9.

network corresponding to different values of m, as obtained by the Meca program, are shown in Fig. 19-12. Observe that the TC plots are highly multivalued and contain separate branches. The sense of the directors on each isocline is determined by using Meca to find either the i_2-vs.-v_1 TC plot or the i_1-vs.-v_1 DP plot of the isocline network shown in Fig. 19-11c. Observe also that

Fig. 19-11. A twin tunnel-diode switching circuit and its associated isocline network 1.



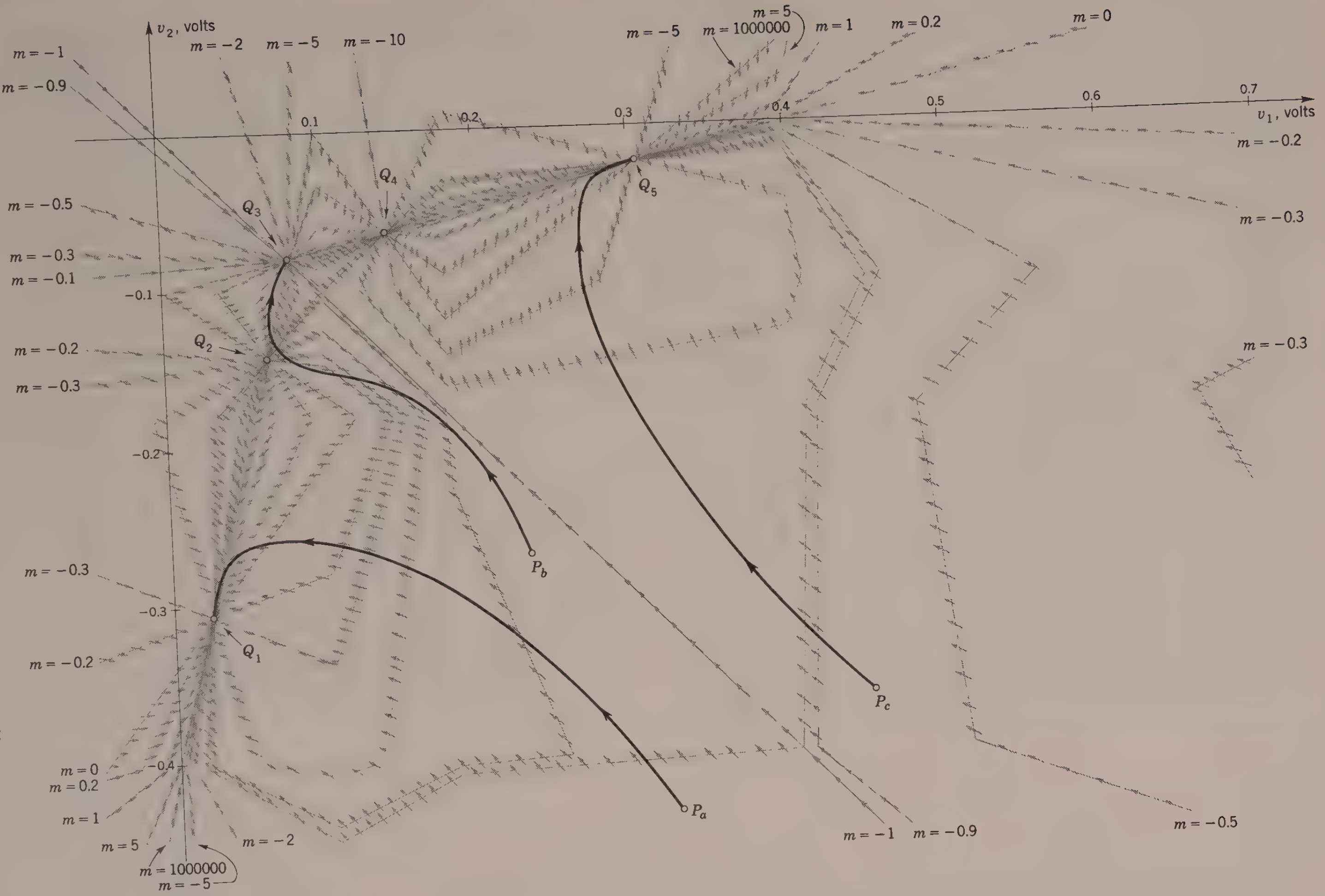


Fig. 19-12. The vector field and trajectories associated with the twin tunnel-diode switching circuit.

the TC plots intersect at the five equilibrium states Q_1, Q_2, Q_3, Q_4, and Q_5. An examination of the vector field near each equilibrium state shows that Q_1, Q_3, and Q_5 are stable, whereas Q_2 and Q_4 are unstable.

Exercise 1: Prove the inductor-loaded isocline network theorem and state the analogous properties corresponding to the $m = 0$ and $m = \infty$ isoclines of the capacitor-loaded network.

Exercise 2: Prove the capacitor-inductor-loaded isocline network theorem and state the analogous properties corresponding to the $m = 0$ and $m = \infty$ isoclines of the capacitor-loaded network.

Exercise 3: Derive the isocline equation for the tunnel-diode switching network shown in Fig. 19-8*a* by using the associated isocline network 2.

Exercise 4: Draw the isocline network 2 associated with the twin tunnel-diode switching circuit shown in Fig. 19-11*a*.

Exercise 5: Show that the direction of the trajectories shown in Fig. 19-9 can also be found from the sign of the voltage across the inductor and the sign of the current in the capacitor at $t = t_0$.

Exercise 6: Explain how to find the direction of the trajectories of a second-order capacitor (inductor)-loaded network as obtained by the isocline network approach.

19-3 FROM QUALITATIVE TO QUANTITATIVE ANALYSIS

The phase-plane technique is useful for obtaining the *qualitative properties* of the solution waveforms. It tells us the number and location of the *equilibrium states* of the network and whether they are stable or unstable. It also tells us the number and location of *limit cycles* of the network. Moreover, it shows how the location of initial states affects the trajectories of the network. In other words, the phase-plane technique may be called a *global analysis method* because it gives a complete picture of a network's behavior.

Once these qualitative properties are found, we may then want to perform a quantitative analysis by finding the exact solution waveform corresponding to some meaningful choice of initial states.[1] For example, one of the most often sought solutions is the waveform corresponding to a limit cycle because it represents the steady-state behavior. In this case the initial state may be chosen to be any point which lies (approximately) on the limit cycle. Another commonly sought solution is the transient waveform during the triggering operation of switching circuits. In this case the important information that is usually desired is the switching

[1] The time-scaling procedure is seldom used for this purpose because simpler and more exact quantitative methods are available.

speed of the network. This can be found by choosing the initial state to coincide with the point on the phase plane for which the triggering signal is first applied. The time it takes the resulting trajectory to "approach" the equilibrium state gives a measure of the switching speed.[1]

Several *quantitative* methods of analysis are available. If the normal-form equation can be easily found in analytic form, then the *numerical techniques* presented in Sec. 4-7 are probably the most straightforward methods to use, with the help of a computer, of course. On the other hand, if the nonlinear devices do not have simple analytical representations, then the piecewise-linear approach to be developed in Sec. 19-5 is usually more practical. As usual, the computer is invariably needed.

19-4 ANALYTICAL SOLUTIONS OF AUTONOMOUS SECOND-ORDER LINEAR NETWORKS

Since the piecewise-linear approach to be developed in the next section is based on the solution of autonomous second-order *linear* networks, let us derive first the analytical solution of this class of networks. The normal-form equations of any autonomous second-order linear network are always of the form

$$\frac{dx_1}{dt} = a_1x_1 + b_1x_2 + c_1 \tag{19-35}$$

$$\frac{dx_2}{dt} = a_2x_1 + b_2x_2 + c_2 \tag{19-36}$$

where the parameters a_i, b_i, and c_i are constants. The equilibrium state of this network is obtained by equating Eqs. (19-35) and (19-36) to zero and solving for x_1 and x_2; namely,

$$x_1 = \left(\frac{b_1c_2 - b_2c_1}{a_1b_2 - a_2b_1}\right) \equiv \gamma_1 \tag{19-37}$$

$$x_2 = \left(\frac{a_2c_1 - a_1c_2}{a_1b_2 - a_2b_1}\right) \equiv \gamma_2 \tag{19-38}$$

The parameters c_1 and c_2 in Eqs. (19-35) and (19-36) are due to the dc independent sources in the network. If the network does not contain any independent source, then $c_1 = c_2 = 0$, $\gamma_1 = \gamma_2 = 0$, and the origin is the only equilibrium state.

---

[1] Although in theory it takes an infinite time for any trajectory to "arrive" at an equilibrium state, it usually takes very little time to "approach" it, i.e., to come within an ϵ-neighborhood of the equilibrium state. This is analogous to the first-order case where an exponential solution is practically in steady state after five time constants.

Since no background in differential equations is assumed in this book, we shall merely state that the solutions (*complete response*) of Eqs. (19-35) and (19-36) corresponding to any initial state $x_1(0)$ and $x_2(0)$ are given by[1]

$$x_1(t) = \alpha_1 e^{\lambda_1 t} + \beta_1 e^{\lambda_2 t} + \gamma_1 \qquad t \geq 0 \tag{19-39}$$

$$x_2(t) = \alpha_2 e^{\lambda_1 t} + \beta_2 e^{\lambda_2 t} + \gamma_2 \qquad t \geq 0 \tag{19-40}$$

where γ_1 and γ_2 are given by Eqs. (19-37) and (19-38), and where α_1, β_1, α_2, β_2 are given by

$$\alpha_1 = \left[\frac{1}{\lambda_1 - \lambda_2}\right][x_1(0)(\lambda_1 - b_2) + x_2(0)b_1 + (c_1 + \gamma_1\lambda_2)] \tag{19-41}$$

$$\beta_1 = -\left[\frac{1}{\lambda_1 - \lambda_2}\right][x_1(0)(\lambda_2 - b_2) + x_2(0)b_1 + (c_1 + \gamma_1\lambda_1)] \tag{19-42}$$

$$\alpha_2 = \left[\frac{1}{\lambda_1 - \lambda_2}\right][x_1(0)a_2 + x_2(0)(\lambda_1 - a_1) + (c_2 + \gamma_2\lambda_2)] \tag{19-43}$$

$$\beta_2 = -\left[\frac{1}{\lambda_1 - \lambda_2}\right][x_1(0)a_2 + x_2(0)(\lambda_2 - a_1) + (c_2 + \gamma_2\lambda_1)] \tag{19-44}$$

The constants λ_1 and λ_2 in these equations are called *natural modes* (also known as normal modes) of the network and are given, respectively, by

$$\lambda_1 = \tfrac{1}{2}(a_1 + b_2) + \tfrac{1}{2}\sqrt{(a_1 + b_2)^2 - 4(a_1 b_2 - a_2 b_1)} \tag{19-45}$$

$$\lambda_2 = \tfrac{1}{2}(a_1 + b_2) - \tfrac{1}{2}\sqrt{(a_1 + b_2)^2 - 4(a_1 b_2 - a_2 b_1)} \tag{19-46}$$

Observe that whereas the constants α_1, β_1, α_2, and β_2 depend on *both* the initial states $(x_1(0), x_2(0))$ and the network parameters a_i, b_i, and c_i, the natural modes λ_1 and λ_2 are *independent* of the initial states.[2]

To prove that Eqs. (19-39) and (19-40) are indeed the solutions, we must do two things. First, we must verify that when $t = 0$, the expressions on the right side of Eqs. (19-39) and (19-40) are equal, respectively, to the given initial state $x_1(0)$ and $x_2(0)$. Second, we must substitute Eqs. (19-39) and (19-40) into Eqs. (19-35) and (19-36) and show that the two sides are identical. Since the above procedures are straightforward, although quite involved algebraically, we shall leave the proof as an exercise.

---

[1] A review of the time-scaling procedure in Sec. 19-2-2 will reveal that if the same initial state is given at a different initial time t_0, then the corresponding solution waveforms will differ from those obtained with $t_0 = 0$ only by a *translation* along the time axis (by t_0 units). Hence, there is no loss of generality for choosing $t_0 = 0$ in our discussions. This time-translation property is *not* valid for nonautonomous networks.

[2] As we shall see shortly, the natural modes completely characterize the behavior of the network.

An examination of Eqs. (19-41) through (19-44) shows that α_1, β_1, α_2, and β_2 are infinite whenever $\lambda_1 = \lambda_2$. Similarly, an examination of Eqs. (19-45) and (19-46) shows that $a_1b_2 - a_2b_1 = 0$ whenever $\lambda_1 = 0$ or $\lambda_2 = 0$. This in turn implies that γ_1 and γ_2 as defined by Eqs. (19-37) and (19-38) are infinite. The solutions given by Eqs. (19-39) and (19-40) are, therefore, undefined under these conditions. Since the correct solution for these special cases must be found by a limiting process, we shall henceforth consider the more general case where the two natural modes are neither identical nor equal to zero.

An examination of Eqs. (19-45) and (19-46) shows that the natural modes are *complex numbers* whenever

$$(a_1 + b_2)^2 < 4(a_1b_2 - a_2b_1) \tag{19-47}$$

In this case it is more convenient to rewrite Eqs. (19-45) and (19-46) into the form[1]

$$\lambda_1 = \tau + j\omega \tag{19-48}$$

$$\lambda_2 = \tau - j\omega \tag{19-49}$$

where

$$\tau \equiv \tfrac{1}{2}(a_1 + b_2) \tag{19-50}$$

$$\omega \equiv \tfrac{1}{2}\sqrt{4(a_1b_2 - a_2b_1) - (a_1 + b_2)^2} \tag{19-51}$$

If we substitute λ_1 and λ_2 from Eqs. (19-48) and (19-49) into Eq. (19-39), we obtain

$$\begin{aligned} x_1(t) &= \alpha_1 e^{(\tau+j\omega)t} + \beta_1 e^{(\tau-j\omega)t} + \gamma_1 \\ &= e^{\tau t}(\alpha_1 e^{j\omega t} + \beta_1 e^{-j\omega t}) + \gamma_1 \end{aligned} \tag{19-52}$$

Since λ_1 and λ_2 are complex numbers, the constants α_1, β_1, α_2, and β_2 will also be complex numbers. If we substitute Eqs. (19-48) and (19-49) into Eqs. (19-41) to (19-44) and simplify, we find that α_1 and β_1 are *complex conjugates* of each other; namely,

$$\alpha_1 = \operatorname{Re}\alpha_1 + j \operatorname{Im}\alpha_1 \tag{19-53}$$

$$\beta_1 = \operatorname{Re}\alpha_1 - j \operatorname{Im}\alpha_1 \tag{19-54}$$

where

$$\operatorname{Re}\alpha_1 \equiv \text{real part of } \alpha_1 = \tfrac{1}{2}[x_1(0) - \gamma_1] \tag{19-55}$$

---

[1] The letter j is used to denote $\sqrt{-1}$. It is chosen in favor of the more conventional notation i in order to avoid any possible confusion with i meaning current.

$$\operatorname{Im}\alpha_1 \equiv \text{imaginary part of } \alpha_1$$
$$= -\left[\frac{x_1(0)(\tau - b_2) + x_2(0)b_1 + (c_1 + \gamma_1\tau)}{2\omega}\right] \tag{19-56}$$

Similarly, we find that α_2 and β_2 are complex conjugates; namely,

$$\alpha_2 = \operatorname{Re}\alpha_2 + j\operatorname{Im}\alpha_2 \tag{19-57}$$

$$\beta_2 = \operatorname{Re}\alpha_2 - j\operatorname{Im}\alpha_2 \tag{19-58}$$

where

$$\operatorname{Re}\alpha_2 = \tfrac{1}{2}[x_2(0) - \gamma_2] \tag{19-59}$$

$$\operatorname{Im}\alpha_2 = -\left[\frac{x_1(0)a_2 + x_2(0)(\tau - a_1) + (c_2 + \gamma_2\tau)}{2\omega}\right] \tag{19-60}$$

Since α_1 and β_1 are complex conjugates, they can be written as

$$\alpha_1 = |\alpha_1|e^{j\varphi_1} \tag{19-61}$$

$$\beta_1 = |\alpha_1|e^{-j\varphi_1} \tag{19-62}$$

where

$$|\alpha_1| \equiv \sqrt{(\operatorname{Re}\alpha_1)^2 + (\operatorname{Im}\alpha_1)^2} \tag{19-63}$$

$$\varphi_1 \equiv \tan^{-1}\left(\frac{\operatorname{Im}\alpha_1}{\operatorname{Re}\alpha_1}\right) \tag{19-64}$$

Substituting Eqs. (19-61) and (19-62) into Eq. (19-52), we obtain

$$\begin{aligned} x_1(t) &= |\alpha_1|e^{\tau t}(e^{j\omega t}e^{j\varphi_1} + e^{-j\omega t}e^{-j\varphi_1}) + \gamma_1 \\ &= 2|\alpha_1|e^{\tau t}\left(\frac{e^{j(\omega t+\varphi_1)} + e^{-j(\omega t+\varphi_1)}}{2}\right) + \gamma_1 \\ &= 2|\alpha_1|e^{\tau t}\cos(\omega t + \varphi_1) + \gamma_1 \end{aligned} \tag{19-65}$$

By a similar procedure Eq. (19-40) can be written into the form

$$x_2(t) = 2|\alpha_2|e^{\tau t}\cos(\omega t + \varphi_2) + \gamma_2 \tag{19-66}$$

where

$$|\alpha_2| \equiv \sqrt{(\operatorname{Re}\alpha_2)^2 + (\operatorname{Im}\alpha_2)^2} \tag{19-67}$$

and

$$\varphi_2 \equiv \tan^{-1}\left(\frac{\operatorname{Im}\alpha_2}{\operatorname{Re}\alpha_2}\right) \tag{19-68}$$

Summarizing, we find that if the natural modes are complex numbers, the solutions can be recast into the following more convenient forms:

$$x_1(t) = \rho_1 e^{\tau t} \cos(\omega t + \varphi_1) + \gamma_1 \qquad t \geq 0 \tag{19-69}$$

$$x_2(t) = \rho_2 e^{\tau t} \cos(\omega t + \varphi_2) + \gamma_2 \qquad t \geq 0 \tag{19-70}$$

where $\rho_1 \equiv 2|\alpha_1|$ and $\rho_2 \equiv 2|\alpha_2|$.

The above results are summarized in Table 19-1 in two columns corresponding to the *aperiodic* and the *spiral* cases. Given the normal-form equations of any network and the prescribed initial states $x_1(0)$ and $x_2(0)$, we can immediately obtain the solution by direct substitution in the appropriate equations in this table. The natures of the solution waveforms corresponding to three typical cases are shown in each column of Table 19-1, together with the associated trajectories. Observe that in the aperiodic case, the trajectory *enters* the stable equilibrium state (case 1) at a unique slope. On the other hand, the trajectory in the spiral case *approaches* the stable equilibrium state (case 1) with an undefined slope because the trajectory is a spiral.

Exercise 1: Prove that the natural modes λ_1 and λ_2 satisfy the following properties:

(*a*) $\lambda_1 + \lambda_2 = (a_1 + b_2)$

(*b*) $\lambda_1 \lambda_2 = a_1 b_2 - a_2 b_1$

Exercise 2: Prove that Eqs. (19-39) and (19-40) are indeed the solutions of Eqs. (19-35) and (19-36).

Exercise 3: Show that if the natural modes are complex numbers, the solutions can be recast into the following equivalent form:

$$x_1(t) = \zeta_1 e^{\tau t} \sin(\omega t + \theta_1) + \gamma_1 \qquad t \geq 0$$

$$x_2(t) = \zeta_2 e^{\tau t} \sin(\omega t + \theta_2) + \gamma_2 \qquad t \geq 0$$

Specify ζ_1 and ζ_2 in terms of the parameters a_i, b_i, c_i and the initial states $x_1(0)$ and $x_2(0)$.

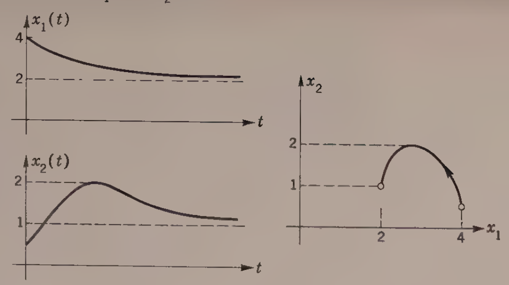
Exercise 4: Specify a set of parameters for the normal-form equations in order to obtain the solution waveforms corresponding to the six cases listed in Table 19-1. Sketch the associated solution waveforms and trajectories.

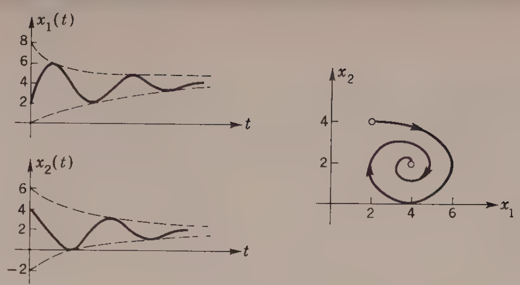
19-5 PIECEWISE-LINEAR ANALYSIS OF AUTONOMOUS SECOND-ORDER NONLINEAR NETWORKS

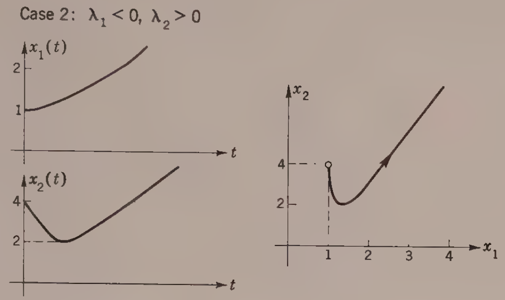
If all nonlinear elements of an autonomous second-order nonlinear network are characterized by piecewise-linear curves, then over some appropriately chosen time interval the network is

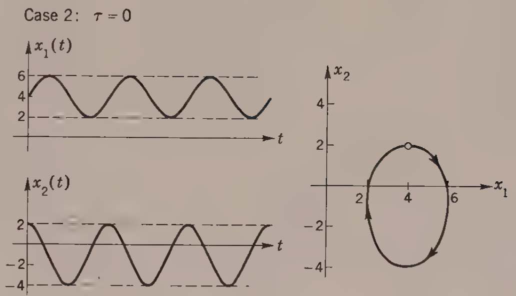
TABLE 19-1 *Summary of solutions of autonomous second-order linear networks and their typical trajectories.*

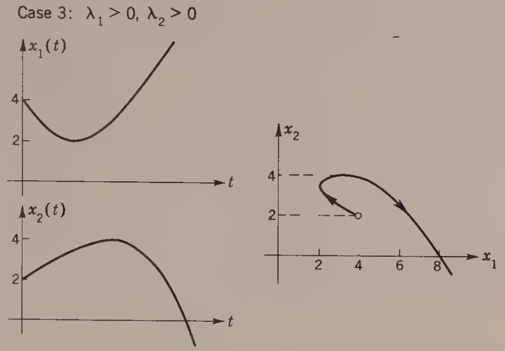
Normal-form equations: $\frac{dx_1}{dt} = a_1x_1 + b_1x_2 + c_1$	$\frac{dx_2}{dt} = a_2x_1 + b_2x_2 + c_2$
Equilibrium states: $x_1 = \gamma_1 \equiv \left(\frac{b_1c_2 - b_2c_1}{a_1b_2 - a_2b_1}\right)$	$x_2 = \gamma_2 \equiv \left(\frac{a_2c_1 - a_1c_2}{a_1b_2 - a_2b_1}\right)$
Aperiodic case: $(a_1 + b_2)^2 > 4(a_1b_2 - a_2b_1)$	Spiral case: $(a_1 + b_2)^2 < 4(a_1b_2 - a_2b_1)$
Natural modes: $\lambda_1 = \frac{1}{2}(a_1 + b_2) + \frac{1}{2}\sqrt{(a_1 + b_2)^2 - 4(a_1b_2 - a_2b_1)}$ $\lambda_2 = \frac{1}{2}(a_1 + b_2) - \frac{1}{2}\sqrt{(a_1 + b_2)^2 - 4(a_1b_2 - a_2b_1)}$ $(\lambda_1 \neq \lambda_2, \lambda_1 \neq 0, \lambda_2 \neq 0)$	Natural modes: $\lambda_1 = \tau + j\omega \qquad \lambda_2 = \tau - j\omega$ $\tau \equiv \frac{1}{2}(a_1 + b_2) \qquad \omega \equiv \frac{1}{2}\sqrt{4(a_1b_2 - a_2b_1) - (a_1 + b_2)^2}$ $(\lambda_1 \neq \lambda_2, \lambda_1 \neq 0, \lambda_2 \neq 0)$
Solutions: $x_1(t) = \alpha_1 e^{\lambda_1 t} + \beta_1 e^{\lambda_2 t} + \gamma_1$ $x_2(t) = \alpha_2 e^{\lambda_1 t} + \beta_2 e^{\lambda_2 t} + \gamma_2$ where $\alpha_1 = \left(\frac{1}{\lambda_1 - \lambda_2}\right)[x_1(0)(\lambda_1 - b_2) + x_2(0)b_1 + (c_1 + \gamma_1\lambda_2)]$ $\beta_1 = -\left(\frac{1}{\lambda_1 - \lambda_2}\right)[x_1(0)(\lambda_2 - b_2) + x_2(0)b_1 + (c_1 + \gamma_1\lambda_1)]$ $\alpha_2 = \left(\frac{1}{\lambda_1 - \lambda_2}\right)[x_1(0)a_2 + x_2(0)(\lambda_1 - a_1) + (c_2 + \gamma_2\lambda_2)]$ $\beta_2 = -\left(\frac{1}{\lambda_1 - \lambda_2}\right)[x_1(0)a_2 + x_2(0)(\lambda_2 - a_1) + (c_2 + \gamma_2\lambda_1)]$	Solutions: $x_1(t) = \rho_1 e^{\tau t} \cos(\omega t + \varphi_1) + \gamma_1$ $x_2(t) = \rho_2 e^{\tau t} \cos(\omega t + \varphi_2) + \gamma_2$ $\rho_1 = 2\sqrt{(\operatorname{Re}\alpha_1)^2 + (\operatorname{Im}\alpha_1)^2} \qquad \varphi_1 = \tan^{-1}\left(\frac{\operatorname{Im}\alpha_1}{\operatorname{Re}\alpha_1}\right)$ where $\operatorname{Re}\alpha_1 = \frac{1}{2}[x_1(0) - \gamma_1]$ $\operatorname{Im}\alpha_1 = -\left[\frac{x_1(0)(\tau - b_2) + x_2(0)b_1 + (c_1 + \gamma_1\tau)}{2\omega}\right]$ $\rho_2 = 2\sqrt{(\operatorname{Re}\alpha_2)^2 + (\operatorname{Im}\alpha_2)^2} \qquad \varphi_2 = \tan^{-1}\left(\frac{\operatorname{Im}\alpha_2}{\operatorname{Re}\alpha_2}\right)$ $\operatorname{Re}\alpha_2 = \frac{1}{2}[x_2(0) - \gamma_2]$ $\operatorname{Im}\alpha_2 = -\left[\frac{x_1(0)a_2 + x_2(0)(\tau - a_1) + (c_2 + \gamma_2\tau)}{2\omega}\right]$

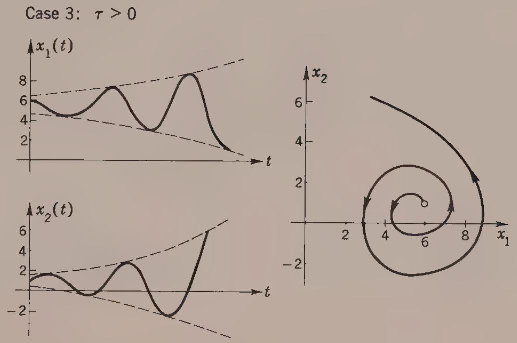

Case 1: $\lambda_1 < 0$, $\lambda_2 < 0$


Case 1: $\tau < 0$


Case 2: $\lambda_1 < 0$, $\lambda_2 > 0$


Case 2: $\tau = 0$


Case 3: $\lambda_1 > 0$, $\lambda_2 > 0$


Case 3: $\tau > 0$

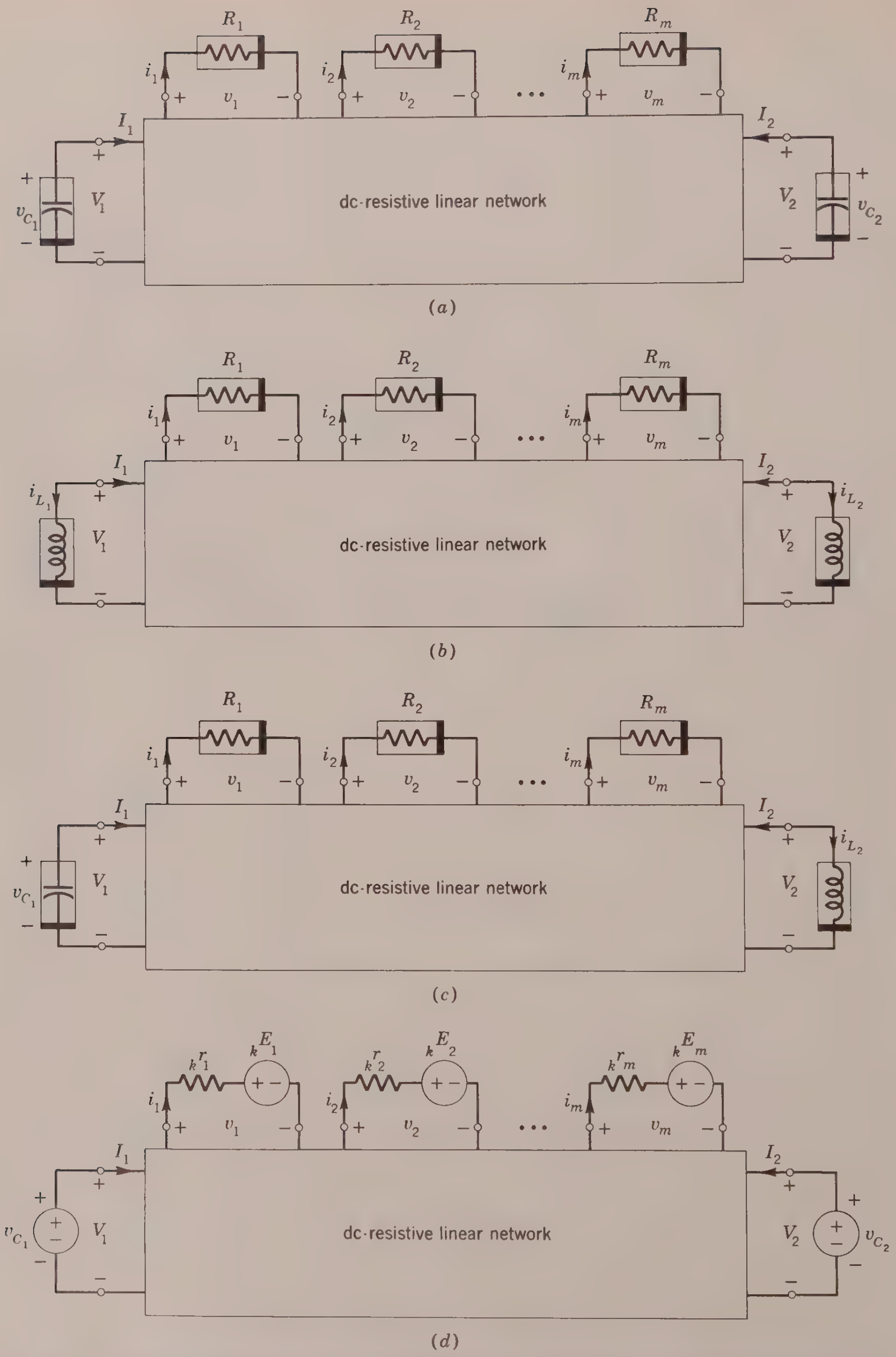
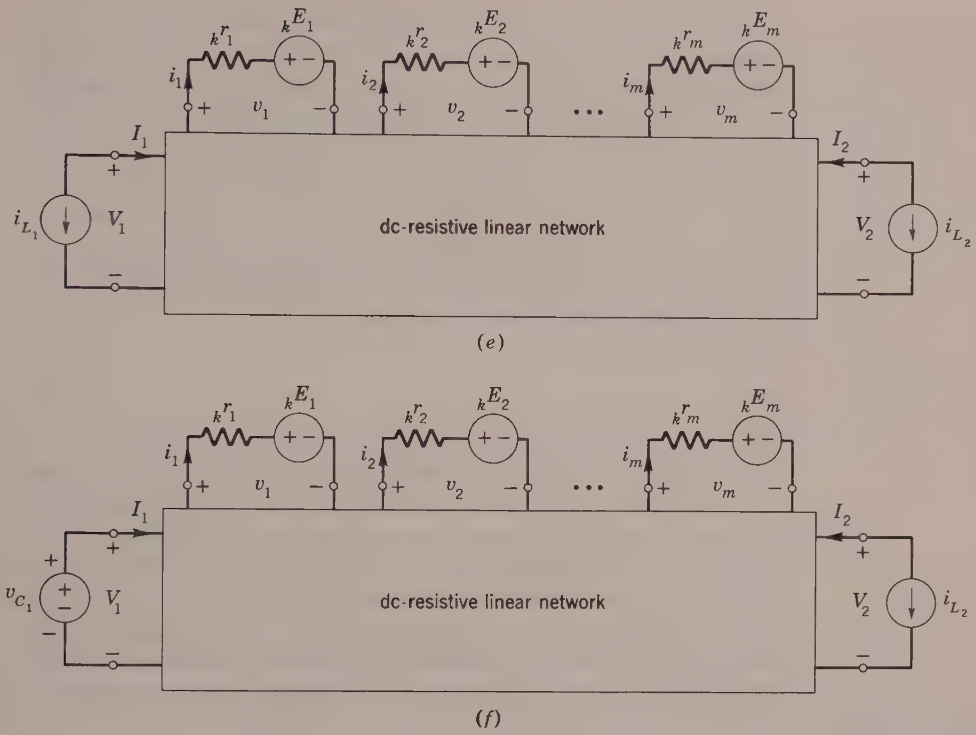


Fig. 19-13. Three autonomous second-order nonlinear networks and their associated iterative piecewise-linear resistive network for determining the switching planes.



(e)

(f)

Fig. 19-13. *(Continued)*

equivalent to a linear network, and its normal-form equations are given by Eqs. (19-35) and (19-36). The analytic solutions tabulated in Table 19-1 can, therefore, be used for obtaining the solutions. But how do we know which parameters a_i, b_i, and c_i to use at any particular time? The key to this problem consists of deriving the *switching plane* associated with the network. The procedure is similar to that developed earlier in Sec. 18-3 for the first-order nonautonomous case.

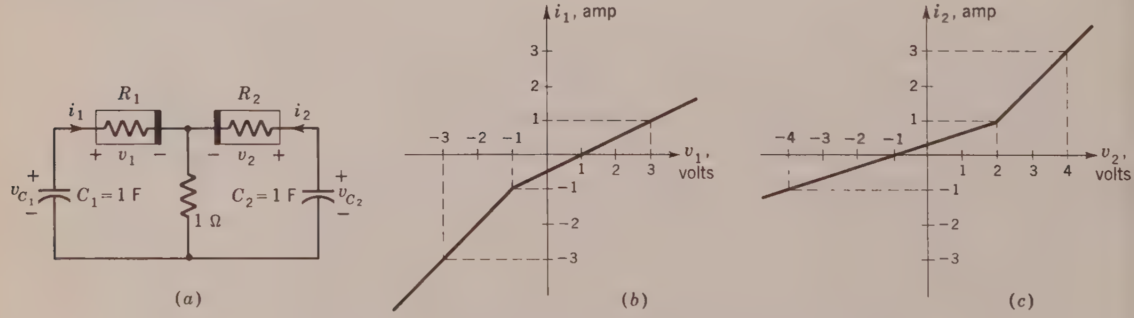
Any autonomous second-order nonlinear network can be represented by one of the three black-box forms shown in Fig. 19-13*a* to *c*. The corresponding iterative piecewise-linear resistive networks used for determining the switching planes are shown in Fig. 19-13*d* to *f*. Observe that each capacitor is replaced by an independent voltage source, and each inductor by an independent current source. Our objective is to solve for the voltage v_j or the current i_j, whichever is more convenient, of *each* nonlinear resistor as a function of the port voltage v_{C_j} or the port current i_{L_j}, depending on whether a capacitor or an inductor is connected across the port.

Once these relationships are obtained, the subsequent procedure is exactly the same as that developed in Sec. 18-3. As usual, we shall assume that the normal form is single-valued so that the switching plane will be subdivided into nonoverlapping "cells." Each cell $(k_1, k_2, \ldots, k_m)$ specifies the region in the phase plane for which resistor R_1 will be operating in segment k_1, resistor R_2 in segment $k_2, \ldots,$ and resistor R_m in segment k_m. Hence, the parameters a_i, b_i, and c_i of the normal-form equations are uniquely defined for each cell in the switching plane. Let us render these ideas concrete by a numerical example.

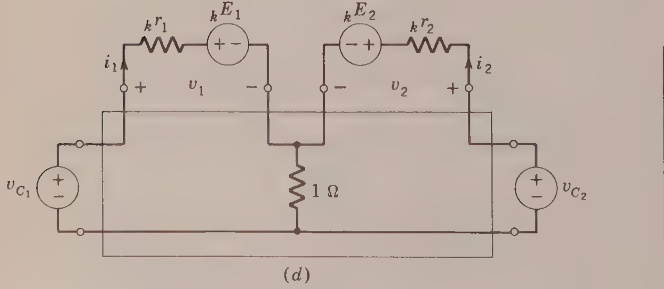
EXAMPLE

Consider the network shown in Fig. 19-14*a*, where the *v*-*i* curves of the nonlinear resistors R_1 and R_2 are shown in Fig. 19-14*b* and *c*, respectively. The iterative piecewise-linear resistive network associated with this network is shown in Fig. 19-14*d*. The pertinent parameters defining the nonlinear resistors are tabulated in Fig. 19-14*e*. The analysis procedure always consists of two parts. The first part consists of the derivation of the switching plane by solving a resistive nonlinear network. The second part consists of obtaining the dynamic solutions corresponding to the pertinent cells.

Fig. 19-14. An autonomous second-order nonlinear network and its associated piecewise-linear resistive network.



(*a*) (*b*) (*c*)



(*d*)

R_j	k	$_k r_j$	$_k E_j$	$_k D_j(v)$	$_k D_j(i)$
R_1	1	1.0	0	$(-\infty, -1.0)$	$(-\infty, -1.0)$
	2	2.0	1.0	$(-1.0, \infty)$	$(-1.0, \infty)$
R_2	1	3.0	−1.0	$(-\infty, 2.0)$	$(-\infty, 1.0)$
	2	1.0	1.0	$(2.0, \infty)$	$(1.0, \infty)$

(*e*)

Part 1: Switching-plane derivation Let us solve for the voltages v_1 and v_2 of the network in Fig. 19-14*d* as a function of v_{C_1} and v_{C_2}.† By straightforward manipulation of the equations of motion, the following transfer-voltage relationships are obtained:

$$v_1=\left[\frac{{}_kr_1(1+{}_kr_2)}{\Delta}\right]v_{C_1}-\left(\frac{{}_kr_1}{\Delta}\right)v_{C_2}+\left\{{}_kE_1+\frac{{}_kr_1[{}_kE_2-(1+{}_kr_2){}_kE_1]}{\Delta}\right\} \tag{19-71}$$

$$v_2=-\left(\frac{{}_kr_2}{\Delta}\right)v_{C_1}+\left[\frac{{}_kr_2(1+{}_kr_1)}{\Delta}\right]v_{C_2}+\left\{{}_kE_2+\frac{{}_kr_2[{}_kE_1-(1+{}_kr_1){}_kE_2]}{\Delta}\right\} \tag{19-72}$$

where

$$\Delta\equiv(1+{}_kr_1)(1+{}_kr_2)-1 \tag{19-73}$$

Since each of the *v-i* curves contains two segments, we must iterate four times. The pertinent equations corresponding to each of the four segment combinations are tabulated in Table 19-2. The equations in the last column of this table specify the *switching lines* separating each cell. They are obtained by replacing the inequality signs in column 4 of this table by equalities and then solving for v_{C_2} as a function of v_{C_1}. For this example each cell is characterized by only two switching lines because one of the two inequality limits in column 4 is $\pm\infty$. In the more general case, each cell is defined by at most $2m$ switching lines, where m is the total number of nonlinear resistors. Also in the general case, each pair of inequalities in column 4 defines a *parallel strip* of points so that the *common region* defining each *cell* is a *convex polygon.* For the present example one side of the convex polygon of each cell is located at infinity. The switching plane can now be easily constructed from Table 19-2 and is shown in Fig. 19-15.

An examination of the above derivation procedures shows that the transfer-voltage relationships for deriving the switching plane depend only on the network topology and not on the *v-i* curves of the nonlinear resistors. Hence, once the transfer voltage (or current) relationships of a particular network are derived, it is a simple matter to obtain the switching plane corresponding to a different set of nonlinear resistors. A general computer program can be easily written for this purpose. The user will simply supply the computer with the transfer relationships and the *v-i* curves of the nonlinear resistors. The computer will automatically calculate the switching lines. In fact, the Meca program described in Sec. 12-5-2 can be used to derive the switching plane automatically; i.e., the user need not even derive the transfer relationships.

---

† Actually, for this particular network it would be simpler to solve for i_1 and i_2 as a function of v_{C_1} and v_{C_2}.

TABLE 19-2 *Summary of equations for deriving the switching plane.*

Segment combination (k_1,k_2)	Transfer-voltage relationship from R_1 $v_1 = {}_kH_{11}(v_{C_1}) + {}_kH_{12}(v_{C_2}) + {}_kK_1$ [Eq. (19-71)]	Transfer-voltage relationship from R_2 $v_2 = {}_kH_{21}(v_{C_1}) + {}_kH_{22}(v_{C_2}) + {}_kK_2$ [Eq. (19-72)]	Inequalities defining each cell (k_1,k_2)	Switching lines defining the boundaries of each cell (k_1,k_2)
(1,1)	$v_1 = \frac{4}{7}v_{C_1} - \frac{1}{7}v_{C_2} - \frac{1}{7}$ $-\infty < v_1 < -1.0$	$v_2 = -\frac{3}{7}v_{C_1} + \frac{6}{7}v_{C_2} - \frac{1}{7}$ $-\infty < v_2 < 2.0$	$-\infty < \frac{4}{7}v_{C_1} - \frac{1}{7}v_{C_2} < -\frac{6}{7}$ $-\infty < -\frac{3}{7}v_{C_1} + \frac{6}{7}v_{C_2} < \frac{15}{7}$	$v_{C_2} = 4v_{C_1} + 6$ $v_{C_2} = 0.5v_{C_1} + 2.5$
(1,2)	$v_1 = \frac{2}{3}v_{C_1} - \frac{1}{3}v_{C_2} + \frac{1}{3}$ $-\infty < v_1 < -1.0$	$v_2 = -\frac{1}{3}v_{C_1} + \frac{2}{3}v_{C_2} + \frac{1}{3}$ $2.0 < v_2 < \infty$	$-\infty < \frac{2}{3}v_{C_1} - \frac{1}{3}v_{C_2} < -\frac{4}{3}$ $\frac{5}{3} < -\frac{1}{3}v_{C_1} + \frac{2}{3}v_{C_2} < \infty$	$v_{C_2} = 2v_{C_1} + 4$ $v_{C_2} = 0.5v_{C_1} + 2.5$
(2,2)	$v_1 = \frac{4}{5}v_{C_1} - \frac{2}{5}v_{C_2} + \frac{3}{5}$ $-1.0 < v_1 < \infty$	$v_2 = -\frac{1}{5}v_{C_1} + \frac{3}{5}v_{C_2} + \frac{3}{5}$ $2.0 < v_2 < \infty$	$-\frac{8}{5} < \frac{4}{5}v_{C_1} - \frac{2}{5}v_{C_2} < \infty$ $\frac{7}{5} < -\frac{1}{5}v_{C_1} + \frac{3}{5}v_{C_2} < \infty$	$v_{C_2} = 2v_{C_1} + 4$ $v_{C_2} = \frac{1}{3}v_{C_1} + \frac{7}{3}$
(2,1)	$v_1 = \frac{8}{11}v_{C_1} - \frac{2}{11}v_{C_2} + \frac{1}{11}$ $-1.0 < v_1 < \infty$	$v_2 = -\frac{3}{11}v_{C_1} + \frac{9}{11}v_{C_2} + \frac{1}{11}$ $-\infty < v_2 < 2.0$	$-\frac{12}{11} < \frac{8}{11}v_{C_1} - \frac{2}{11}v_{C_2} < \infty$ $-\infty < -\frac{3}{11}v_{C_1} + \frac{9}{11}v_{C_2} < \frac{21}{11}$	$v_{C_2} = 4v_{C_1} + 6$ $v_{C_2} = \frac{1}{3}v_{C_1} + \frac{7}{3}$

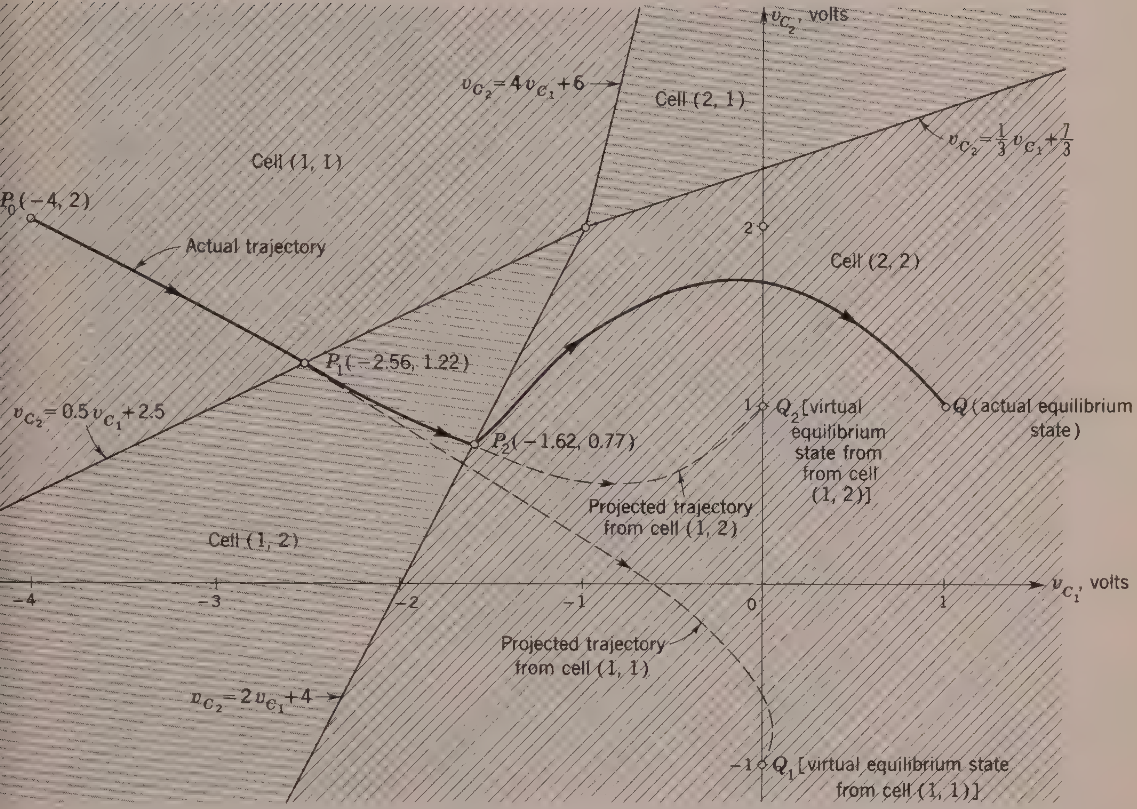


Fig. 19-15. The switching plane for the network shown in Fig. 19-14*a* and a typical trajectory.

Part 2: Deriving the dynamic solutions The normal-form equations are derived from the iterative piecewise-linear network in Fig. 19-14*d* upon replacing the two voltage sources by the original capacitors.[1] The results are easily found to be given by

$$\frac{dv_{C_1}}{dt} = -\frac{1}{C_1}\left\{\left[\frac{(1+{}_kr_2)}{\Delta}\right]v_{C_1} - \left(\frac{1}{\Delta}\right)v_{C_2} + \left[\frac{{}_kE_2-(1+{}_kr_2){}_kE_1}{\Delta}\right]\right\} \tag{19-74}$$

$$\frac{dv_{C_2}}{dt} = -\frac{1}{C_2}\left\{\left(-\frac{1}{\Delta}\right)v_{C_1} + \left[\frac{(1+{}_kr_1)}{\Delta}\right]v_{C_2} + \left[\frac{{}_kE_1-(1+{}_kr_1){}_kE_2}{\Delta}\right]\right\} \tag{19-75}$$

where Δ is defined by Eq. (19-73). Comparing these equations with Eqs. (19-35) and (19-36), we identify $x_1 = v_{C_1}$ and $x_2 = v_{C_2}$. The corresponding expressions for the parameters a_i, b_i, and c_i are tabulated in column 1 of Table 19-3. Substituting the pertinent values of ${}_kr_j$ and ${}_kE_j$ from Fig. 19-14*e*, we obtain the numerical values of a_i, b_i, and c_i for each of the four cells as tabulated in columns 2 to 5 of Table 19-3.

---

[1] If the capacitor is nonlinear, it can be approximated by a piecewise-linear v-q curve, and the same method is applicable with minor modifications. A dual remark applies for nonlinear inductors.

It is now a simple matter to find the solution waveforms $v_{C_1}(t)$ and $v_{C_2}(t)$ corresponding to any initial state $[v_{C_1}(0),v_{C_2}(0)]$. For example, suppose the desired initial state is given by $v_{C_1}(0) = -4$ volts and $v_{C_2}(0) = 2$ volts. This corresponds to point P_0 in Fig. 19-15. The solutions $v_{C_1}(t)$ and $v_{C_2}(t)$ can be easily obtained with the help of Table 19-1. The numerical values of the pertinent parameters for the general solutions in this table are tabulated in Table 19-4 for each of the four cells. Observe that since $(a_1 + b_2)^2 > 4(a_1b_2 - a_2b_1)$ in all cases, the parameters must be evaluated for the *aperiodic case.*

Since the initial state P_0 falls on cell (1,1), we must substitute the parameters in column 2 of Table 19-3 with the initial state $x_1(0) = -4$ and $x_2(0) = 2$ into the pertinent equations in Table 19-1. The results are given by

$$v_{C_1}(t) = 0.525e^{-0.228t} - 4.525e^{-0.628t} \qquad 0 \le t \le t_1 \tag{19-76}$$

$$v_{C_2}(t) = 1.143e^{-0.228t} + 1.857e^{-0.628t} - 1 \qquad 0 \le t \le t_1 \tag{19-77}$$

The trajectory corresponding to Eqs. (19-76) and (19-77) is plotted in Fig. 19-15, where the portion outside cell (1,1) is shown by dotted lines for comparison purposes. Observe that had the param-

TABLE 19-3 *The pertinent parameters defining the normal-form equations for each cell (k_1,k_2).*

Parameters for normal-form equations	Cell (1,1)	Cell (1,2)	Cell (2,2)	Cell (2,1)
$a_1 = -\dfrac{1}{C_1}\left(\dfrac{1 + {}_kr_2}{\Delta}\right)$	$-4/7$	$-2/3$	$-2/5$	$-4/11$
$b_1 = \dfrac{1}{C_1}\left(\dfrac{1}{\Delta}\right)$	$1/7$	$1/3$	$1/5$	$1/11$
$c_1 = -\dfrac{1}{C_1}\left[\dfrac{{}_kE_2 - (1 + {}_kr_2){}_kE_1}{\Delta}\right]$	$1/7$	$-1/3$	$1/5$	$5/11$
$a_2 = \dfrac{1}{C_2}\left(\dfrac{1}{\Delta}\right)$	$1/7$	$1/3$	$1/5$	$1/11$
$b_2 = -\dfrac{1}{C_2}\left(\dfrac{1 + {}_kr_1}{\Delta}\right)$	$-2/7$	$-2/3$	$-3/5$	$-3/11$
$c_2 = -\dfrac{1}{C_2}\left[\dfrac{{}_kE_1 - (1 + {}_kr_1){}_kE_2}{\Delta}\right]$	$-2/7$	$2/3$	$2/5$	$-4/11$

TABLE 19-4 *The pertinent parameters defining the solutions for each cell (k_1,k_2).*

Pertinent parameters	Cell (1,1)	Cell (1,2)	Cell (2,2)	Cell (2,1)
$(a_1 + b_2)^2$	0.73	1.78	1.0	0.405
$4(a_1b_2 - a_2b_1)$	0.57	1.33	0.80	0.364
Equilibrium state (γ_1,γ_2)	$(0,-1)$	$(0,1)$	$(1,1)$	$(1,-1)$
Natural mode λ_1	-0.228	-0.333	-0.277	-0.218
Natural mode λ_2	-0.628	-1.0	-0.723	-0.418

eters remained unchanged, this trajectory would arrive at the "virtual" equilibrium state $Q_1(0,-1)$.† Of course, this trajectory is valid only up to the point $P_1(-2.56,1.22)$ where it intersects a *switching line.* The coordinates at P_1 can be obtained either graphically as shown in Fig. 19-15 or by substituting Eqs. (19-76) and (19-77) for v_{C_1} and v_{C_2} into the switching-line equation $v_{C_2} = 0.5v_{C_1} + 2.5$; namely,

$$1.143e^{-0.228t} + 1.857e^{-0.628t} - 1 = 0.2625e^{-0.228t} - 2.2625e^{-0.628t} + 2.5$$

Upon simplification, we obtain

$$0.8805e^{-0.228t} + 4.1195e^{-0.628t} = 3.5 \qquad (19\text{-}78)$$

Equation (19-78) is a *nonlinear transcendental equation* in t and can be solved by either numerical or graphical techniques. The solution is found to be given approximately by $t = 0.65 \equiv t_1$. The coordinates are found by substituting $t = 0.65$ into Eqs. (19-76) and (19-77); namely, $v_{C_1}(t = 0.65) = -2.56$ and $v_{C_2}(t = 0.65) = 1.22$.

At $t = 0.65$ the trajectory enters cell (1,2), and hence we must now substitute the parameters in column 3 of Table 19-3 with the new initial state $x_1(0) = -2.56$ and $x_2(0) = 1.22$ into the pertinent equations in Table 19-1 to obtain

$$v_{C_1}(t) = -1.17e^{-0.333(t-t_1)} - 1.39e^{-(t-t_1)} \qquad t_1 \le t \le t_2 \qquad (19\text{-}79)$$

$$v_{C_2}(t) = -1.17e^{-0.333(t-t_1)} + 1.39e^{-(t-t_1)} + 1 \qquad t_1 \le t \le t_2 \qquad (19\text{-}80)$$

Observe that since the initial time for this case is t_1 and not zero, the solution must consist of a translation of the corresponding solution in Table 19-1 by t_1 units along the time axis. This accounts for the term $(t - t_1)$ in the exponent of the exponentials. Another reason for inserting t_1 in the exponents is so that when $t = t_1$, v_{C_1} and v_{C_2} will satisfy the *new* initial state at t_1 and not at $t = 0$. The trajectory corresponding to Eqs. (19-79) and (19-80) is plotted in Fig. 19-15. Again, observe that it is valid only up to the point P_2 where it intersects with the switching line $v_{C_2} = 2v_{C_1} + 4$. Had the parameters for cell (1,2) remained unchanged, this trajectory (shown dotted) would have continued until it arrived at the virtual equilibrium state $Q_2(0,1)$. The time it takes this trajectory to arrive at P_2 is found to be equal to $t = 1.35 \equiv t_2$. Hence, it takes $\Delta t = 0.7$ units of time for the trajectory to move from P_1 to P_2. The coordinates at P_2 are $v_{C_1}(t_2) = -1.62$ and $v_{C_2}(t_2) = 0.77$.

---

† Any equilibrium state (γ_1,γ_2) which falls outside the cell it pertains to is said to be a *virtual equilibrium state.*

At $t = t_2$ the trajectory enters cell (2,2), and hence we must substitute the parameters in column 4 of Table 19-3 with the new initial state $x_1(0) = -1.62$ and $x_2(0) = 0.77$ into the pertinent equations in Table 19-1 to obtain

$$v_{C_1}(t) = -2.0e^{-0.277(t-t_2)} - 0.62e^{-0.723(t-t_2)} + 1 \qquad t_2 \leq t < \infty \tag{19-81}$$

$$v_{C_2}(t) = 2.19e^{-0.277(t-t_2)} - 2.42e^{-0.723(t-t_2)} + 1 \qquad t_2 \leq t < \infty \tag{19-82}$$

The trajectory corresponding to Eqs. (19-81) and (19-82) is shown in Fig. 19-15. Observe that this time, the trajectory is valid for all times $t \geq t_2$ and will eventually arrive at the equilibrium state $v_{C_1} = 1$ and $v_{C_2} = 1$.

To summarize, we found that the piecewise-linear method consists of two parts. Part 1 consists of dividing the phase plane into nonoverlapping cells. Part 2 consists of finding the solutions with the help of the general expressions derived in Table 19-1 and the switching plane. The most time-consuming task in this part consists of finding the exact time and coordinates for which a trajectory crosses a switching line. Analytically, this involves solving a nonlinear transcendental equation of the general form

$$Ae^{\lambda_1 t} + Be^{\lambda_2 t} = K$$

To simplify the task, this equation can be easily solved by a computer or by the construction of tables and normalized graphs.

It should now be clear that the algorithmic nature of this approach is ideally suited for a completely automatic computer solution. This approach is especially useful if the same network is to be analyzed many times with different initial states or different capacitance or inductance values. Then, the switching plane remains unchanged, and only part 2 need be carried out.

Exercise 1: Derive the switching plane of the network shown in Fig. 19-14*d* from the relationships i_1 and i_2 as a function of v_{C_1} and v_{C_2}.

Exercise 2: Derive the switching plane of the network shown in Fig. 19-14*d* from the relationships v_1 and i_2 as a function of v_{C_1} and v_{C_2}.

Exercise 3: Derive the switching plane for the network shown in Fig. 19-14*a* by choosing the iterative Norton equivalent circuit for R_1 and R_2.

Exercise 4: Derive the switching plane for the network shown in Fig. 19-14*a* but with the capacitors replaced by two 1-H inductors.

Exercise 5: Derive the switching plane for the network shown in Fig. 19-14*a* but with the capacitor C_2 replaced by a 1-H inductor.

Exercise 6: Find the solution $v_{C_1}(t)$ and $v_{C_2}(t)$ of the network shown in Fig. 19-14*a* with $C_1 = 1$ F and $C_2 = -1$ F. Assume that $v_{C_1}(0) = -4$ volts and $v_{C_2}(0) = 2$ volts.

Exercise 7: Repeat Exercise 6 with $C_1 = C_2 = -1$ F.

Exercise 8: Repeat Exercise 6 with the initial state $v_{C_1}(0) = -1.0$ volt and $v_{C_2}(0) = 3.0$ volts.

19-6 BEHAVIOR OF TRAJECTORIES NEAR AN EQUILIBRIUM STATE

It was mentioned in Sec. 13-6 that the behavior of autonomous nonlinear networks can often be determined by studying the behavior of trajectories in the vicinity of each *equilibrium state* of the network.[1] This is especially true for second-order networks because of the many geometrical interpretations that we are about to derive. Not only can we determine the stability of an equilibrium state for this class of networks, but we can also sketch the trajectories around each singular point without actually obtaining the solution.

For simplicity, we shall assume that the network is piecewise-linear and that the singular points do not lie on a switching line. Hence, the normal-form equations that describe the behavior of the trajectories near each singular point (γ_1,γ_2) are given by Eqs. (19-35) and (19-36), which we reproduce here for convenience.

$$\frac{dx_1}{dt} = a_1x_1 + b_1x_2 + c_1 \tag{19-83}$$

$$\frac{dx_2}{dt} = a_2x_1 + b_2x_2 + c_2 \tag{19-84}$$

Since our objective is to study the properties of the trajectories near a singular point $Q(\gamma_1,\gamma_2)$, such as the shaded region shown in Fig. 19-16*a*, it is convenient for us to translate the origin of our coordinate system to point Q itself. This is easily done by defining

$$x_1 = y_1 + \gamma_1 \tag{19-85}$$

$$x_2 = y_2 + \gamma_2 \tag{19-86}$$

Observe that the singular point $x_1 = \gamma_1$, $x_2 = \gamma_2$ becomes $y_1 = 0$, $y_2 = 0$ in the new coordinate system as shown in Fig. 19-16*b*.

---

[1] In more mathematically oriented books, the term *singular point* is usually used in place of *equilibrium state*. Since they are synonyms, we shall use them interchangeably in order that the reader will not develop any prejudice over either term.

In terms of the new coordinates y_1 and y_2, Eqs. (19-83) and (19-84) reduce to

$$\frac{dy_1}{dt} = a_1y_1 + b_1y_2 \tag{19-87}$$

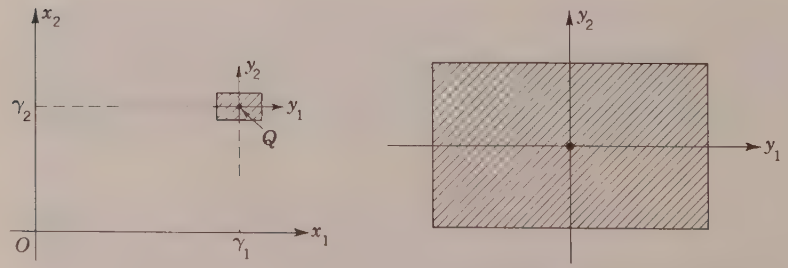
$$\frac{dy_2}{dt} = a_2y_1 + b_2y_2 \tag{19-88}$$

As we expected, the origin $y_1 = 0$, $y_2 = 0$ is the singular point of Eqs. (19-87) and (19-88). The trajectories of this system are the solution of the phase-plane equation

$$\frac{dy_2}{dy_1} = \frac{a_2y_1 + b_2y_2}{a_1y_1 + b_1y_2} \tag{19-89}$$

Since y_1 and y_2 differ only from x_1 and x_2 by a constant [Eqs. (19-85) and (19-86)], any property concerning the trajectories of Eq. (19-89) near the origin remains valid in the vicinity of the singular point $Q(\gamma_1,\gamma_2)$ of Eqs. (19-83) and (19-84).

A singular point is stable if all trajectories nearby tend to it as $t \to \infty$. It is unstable if at least one trajectory nearby moves away from it as $t \to \infty$. The two trajectories shown in Fig. 19-17*a* and *b* represent two typical examples corresponding to the *aperiodic* case in Table 19-1. Similarly, the two trajectories shown in Fig. 19-17*c* and *d* are typical of the *spiral* case. In all cases the origin is assumed to be the singular point, and the arrow points in the direction of increasing time. Hence if P_0 is chosen to be the initial state, then the trajectories in Fig. 19-17*a* and *c* will tend to the origin as $t \to \infty$. The opposite behavior occurs in Fig. 19-17*b* and *d* where the trajectories move away from the origin as $t \to \infty$. Ob-



(*a*) (*b*)

Fig. 19-16. In order to investigate the behavior of the trajectories near a singular point, it is convenient to translate the singular point to the origin.

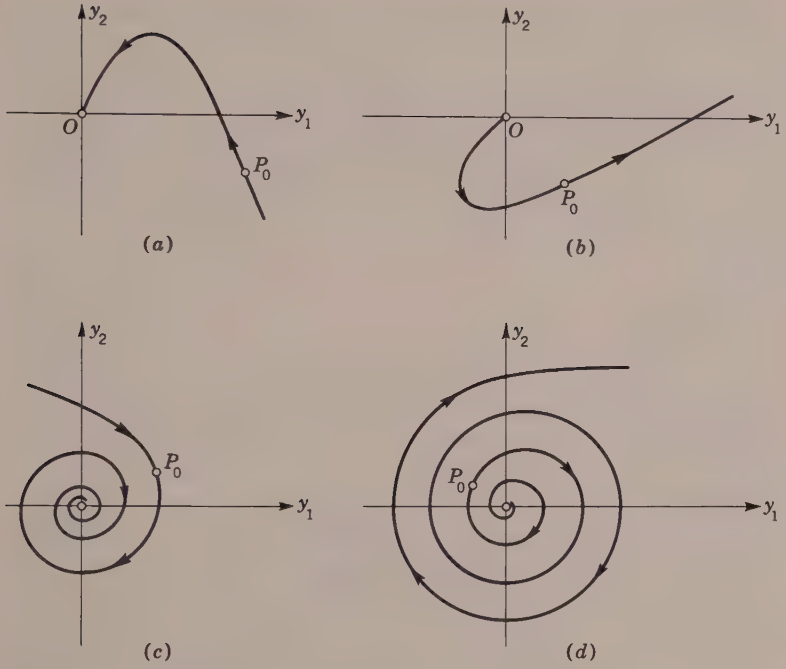


Fig. 19-17. Four representative types of trajectories corresponding to the aperiodic case and the spiral case.

serve, however, that this implies that the trajectories in Figs. 19-17*b* and *d* will tend toward the singular point at $t \to -\infty$.† Hence, we may conclude that a trajectory will tend to the origin either as $t \to \infty$ or as $t \to -\infty$. With this property in mind, let us now turn to a detailed investigation of the properties of trajectories corresponding to the aperiodic case and the spiral case.

19-6-1 LOCAL BEHAVIOR OF TRAJECTORIES IN THE APERIODIC CASE

An examination of the form of solutions of $x_1(t)$ and $x_2(t)$ in the left column of Table 19-1, as well as of the typical trajectories shown in Fig. 19-17*a* and *b*, shows that any trajectory in the aperiodic case will approach the singular point with a unique slope either as $t \to \infty$ or as $t \to -\infty$. Our immediate objective is to derive an expression giving the *slope of approach* S in terms of the parameters a_1, b_1, a_2, and b_2, where S is defined to be

$$S \equiv \lim_{t \to \pm\infty} \frac{y_2(t)}{y_1(t)} \tag{19-90}$$

---

† There is nothing mysterious about "negative time." If we arbitrarily assume the time parameter at P_0 to be $t = 0$, representing the *time now,* then the location of the point on the trajectory 5 sec ago must correspond to $t = -5$. Observe that we can arbitrarily pick the initial point and the initial time because the system is *autonomous.* Essentially, this means that what happened yesterday can be repeated today using the same system with the same initial condition.

By the Lhopital rule, Eq. (19-90) can be written as

$$S = \lim_{t\to\pm\infty}\left[\frac{dy_2(t)/dt}{dy_1(t)/dt}\right] = \lim_{t\to\pm\infty}\left(\frac{a_2y_1 + b_2y_2}{a_1y_1 + b_1y_2}\right)$$

$$= \lim_{t\to\pm\infty}\left[\frac{a_2 + b_2y_2(t)/y_1(t)}{a_1 + b_1y_2(t)/y_1(t)}\right] = \left\{\frac{a_2 + b_2 \lim\limits_{t\to\pm\infty}[y_2(t)/y_1(t)]}{a_1 + b_1 \lim\limits_{t\to\pm\infty}[y_2(t)/y_1(t)]}\right\}$$

or

$$S = \frac{a_2 + b_2S}{a_1 + b_1S} \tag{19-91}$$

Simplifying and solving for S, we find that S may assume two distinct values, S_1 and S_2, where

$$S_1 = \frac{-(a_1 - b_2) + \sqrt{(a_1 - b_2)^2 + 4b_1a_2}}{2b_1} \tag{19-92}$$

$$S_2 = \frac{-(a_1 - b_2) - \sqrt{(a_1 - b_2)^2 + 4b_1a_2}}{2b_1} \tag{19-93}$$

It can be easily verified that the following simple relationships exist between S_1, S_2, and the natural modes λ_1, λ_2 defined in Table 19-1:

$$S_1 = \frac{\lambda_1 - a_1}{b_1} \tag{19-94}$$

$$S_2 = \frac{\lambda_2 - a_1}{b_1} \tag{19-95}$$

We shall now show that the two straight lines

$$y_2 = S_1y_1 \tag{19-96}$$

$$y_2 = S_2y_1 \tag{19-97}$$

are actually trajectories of the network.[1] To prove this, it suffices to show that Eqs. (19-96) and (19-97) are solutions of Eq. (19-89). Hence, substituting Eq. (19-96) into Eq. (19-89), we obtain

$$S_1 = \frac{(a_2 + b_2S_1)}{(a_1 + b_1S_1)} \tag{19-98}$$

---

[1] These must also be the *only* straight-line trajectories because any other straight-line trajectory would have a slope different from Eqs. (19-94) and 19-95). But this contradicts the property that all trajectories must approach the origin with slope S_1 or S_2 as $t \to \pm\infty$.

which coincides with Eq. (19-91). This means that we have an identity, and the result is proved. A similar proof applies to Eq. (19-97).

The solution $y_1(t)$ and $y_2(t)$ corresponding to any trajectory of Eq. (19-89) can be found by substituting $x_1(t)$ and $x_2(t)$ from Table 19-1 into Eqs. (19-85) and (19-86); namely,

$$y_1(t) = \alpha_1 e^{\lambda_1 t} + \beta_1 e^{\lambda_2 t} \tag{19-99}$$

$$y_2(t) = \alpha_2 e^{\lambda_1 t} + \beta_2 e^{\lambda_2 t} \tag{19-100}$$

where λ_1 and λ_2 are the natural modes defined in Table 19-1. Observe that *by definition,* λ_1 *is associated with the "positive" radical sign.* Therefore, for the aperiodic case we have $\lambda_1 > \lambda_2$.

We shall now show that the relative signs of the natural modes determine whether a singular point is stable or unstable. If both $\lambda_1 < 0$ and $\lambda_2 < 0$, then $y_1(t) \to 0$ and $y_2(t) \to 0$, as $t \to \infty$. For the opposite case where both $\lambda_1 > 0$ and $\lambda_2 > 0$, we find that $y_1(t) \to \infty$ and $y_2(t) \to \infty$, as $t \to \infty$. Finally, if $\lambda_1 > 0$, $\lambda_2 < 0$, $\alpha_1 \neq 0$, and $\alpha_2 \neq 0$, then $y_1(t) \to \infty$ and $y_2(t) \to \infty$, as $t \to \infty$.† The above results can be summarized by the following:

STABILITY CRITERIA FOR APERIODIC CASE

If both natural modes are negative real numbers, then the origin is a stable singular point. If at least one natural mode is a positive real number, then the origin is an unstable singular point.

We can immediately draw *four* straight-line trajectories of Eq. (19-89), two of which tend to the origin with slope S_1 and two with slope S_2. The four trajectories corresponding to each of the three sign combinations of λ_1 and λ_2 are shown in Fig. 19-18. Each of these trajectories is called an *eigenvector.*

We found earlier that the eigenvectors are the only straight-line trajectories. All other trajectories must, therefore, be curves. We shall now develop a simple technique for sketching the remaining trajectories. Suppose we divide Eq. (19-100) by Eq. (19-99).

$$\frac{y_2(t)}{y_1(t)} = \frac{\alpha_2 e^{\lambda_1 t} + \beta_2 e^{\lambda_2 t}}{\alpha_1 e^{\lambda_1 t} + \beta_1 e^{\lambda_2 t}} = \frac{\alpha_2 + \beta_2 e^{(\lambda_2 - \lambda_1)t}}{\alpha_1 + \beta_1 e^{(\lambda_2 - \lambda_1)t}} \tag{19-101}$$

---

† The combination $\lambda_1 < 0$ and $\lambda_2 > 0$ does not exist because by our definition $\lambda_1 > \lambda_2$.

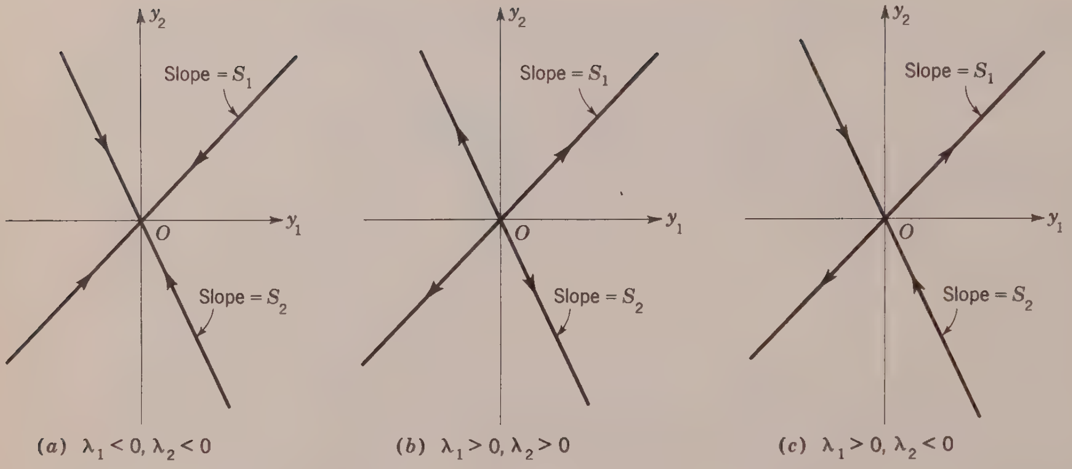

y_2
Slope = S_1
y_1
O
Slope = S_2
(a) $\lambda_1 < 0, \lambda_2 < 0$
(b) $\lambda_1 > 0, \lambda_2 > 0$
(c) $\lambda_1 > 0, \lambda_2 < 0$

Fig. 19-18. There are always four straight-line trajectories passing through each singular point in the aperiodic case. The direction of motion along each trajectory is determined by the relative signs of the natural modes.

Since $\lambda_1 > \lambda_2$ or $(\lambda_2 - \lambda_1) < 0$, the coefficient of t is a negative number. It follows, therefore, from Eq. (19-101) that

$$\lim_{t\to\infty} \frac{y_2(t)}{y_1(t)} = \frac{\alpha_2}{\alpha_1} \tag{19-102}$$

and

$$\lim_{t\to-\infty} \frac{y_2(t)}{y_1(t)} = \frac{\beta_2}{\beta_1} \tag{19-103}$$

Observe that the left side of Eqs. (19-102) and (19-103) have been defined earlier by Eq. (19-90) to be the slope of approach S. Since we have already proved that S can assume only one of two values as given by Eqs. (19-92) and (19-93), it follows that α_2/α_1 and β_2/β_1 must be equal, respectively, to either S_1 or S_2. Substituting the expressions for α_1, α_2, β_1, and β_2 from Table 19-1 into Eqs. (19-102) and (19-103), we obtain

$$\frac{\alpha_2}{\alpha_1} = \frac{x_1(0)a_2 + x_2(0)(\lambda_1 - a_1) + (c_2 + \gamma_2\lambda_2)}{x_1(0)(\lambda_1 - b_2) + x_2(0)b_1 + (c_1 + \gamma_1\lambda_2)} \tag{19-104}$$

$$\frac{\beta_2}{\beta_1} = \frac{x_1(0)a_2 + x_2(0)(\lambda_2 - a_1) + (c_2 + \gamma_2\lambda_1)}{x_1(0)(\lambda_2 - b_2) + x_2(0)b_1 + (c_1 + \gamma_1\lambda_1)} \tag{19-105}$$

Since Eqs. (19-104) and (19-105) are functions of the initial state $[x_1(0),x_2(0)]$, which can be arbitrarily chosen, it is a most surprising property that both α_2/α_1 and β_2/β_1 are equal to constants *independent* of $x_1(0)$ and $x_2(0)$! In view of this property, we may calculate α_2/α_1 and β_2/β_1 by choosing some convenient parameters and initial conditions. One simple choice is $x_1(0) = 0$, $x_2(0) = 1$, $c_1 = 0$, and $c_2 = 0$.† In this case, Eqs. (19-104) and (19-105) reduce to

$$\frac{\alpha_2}{\alpha_1} = \frac{\lambda_1 - a_1}{b_1} \tag{19-106}$$

$$\frac{\beta_2}{\beta_1} = \frac{\lambda_2 - a_1}{b_1} \tag{19-107}$$

Comparing Eqs. (19-106) and (19-107) with Eqs. (19-94) and (19-95), we find

$$\lim_{t\to\infty} \frac{y_2(t)}{y_1(t)} = \frac{\alpha_2}{\alpha_1} = \frac{\lambda_1 - a_1}{b_1} = S_1 \tag{19-108}$$

$$\lim_{t\to-\infty} \frac{y_2(t)}{y_1(t)} = \frac{\beta_2}{\beta_1} = \frac{\lambda_2 - a_1}{b_1} = S_2 \tag{19-109}$$

The interpretation of Eqs. (19-108) and (19-109) are important enough to be stated in the form of a theorem.

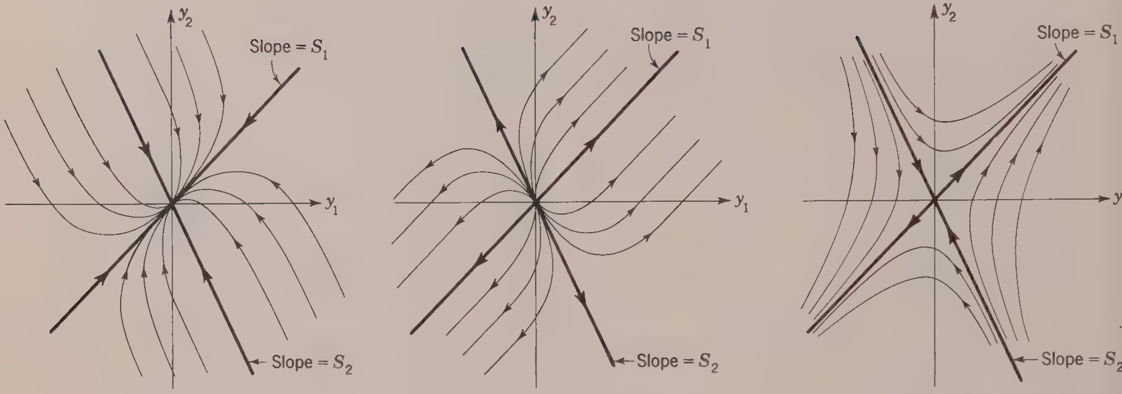
TRAJECTORY SLOPE THEOREM

The slope of any trajectory of Eq. (19-89) must assume the slope $S = S_1$ as $t \to \infty$ and the slope $S = S_2$ as $t \to -\infty$.

Since the slope of the eigenvectors is equal to either S_1 or S_2, it follows from the above theorem that the trajectories must be either parallel or tangent to the eigenvector with slope equal to S_1 as $t \to \infty$. Similarly, the trajectories must be either parallel or tangent to the eigenvector with slope equal to S_2 as $t \to -\infty$. This observation can now be used to sketch the trajectories near each singular point. Let us consider some typical cases.

Case 1: Stable node ($\lambda_1 < 0, \lambda_2 < 0$) Let us arbitrarily assume that the eigenvectors have been drawn (by the simple technique developed above) as in Fig. 19-18*a*. Since the origin is a stable singular point, all trajectories must approach this point as $t \to \infty$. Moreover, each trajectory must be tangent to the eigenvector with slope S_1 as shown in Fig. 19-19*a*. An examination of that portion of each trajectory corresponding to large *negative* time ($t \to -\infty$) shows that they are all drawn approximately parallel to the eigen-

† From Table 19-1 we find that $c_1 = c_2 = 0$ implies $\gamma_1 = \gamma_2 = 0$.



(*a*) $\lambda_1 < 0$, $\lambda_2 < 0$
(stable node)

(*b*) $\lambda_1 > 0$, $\lambda_2 > 0$
(unstable node)

(*c*) $\lambda_1 > 0$, $\lambda_2 < 0$
(saddle point)

Fig. 19-19. The trajectories around each type of singular point can be easily sketched with the help of the trajectory slope theorem.

vector with slope S_2. Since all trajectories in this case are seen to be converging toward a common node, the origin in this case is called a *stable node.* It is important to realize that although the exact shape of the trajectories around a stable node will depend on the parameters of the network, the general behavior, that of converging toward the node, will remain unchanged so long as $\lambda_1 < 0$ and $\lambda_2 < 0$. If we imagine that the trajectories in Fig. 19-19*a* were drawn on top of a transparent rubber sheet, then the new trajectories arising from changing the network parameters can be simulated by *stretching* the rubber sheet, thereby distorting the trajectories, in such a way that the sheet is not punctured by the stretching process.

Case 2: Unstable node ($\lambda_1 > 0$, $\lambda_2 > 0$) Using the same eigenvectors in Fig. 19-18*b*, we now observe that the trajectories must be drawn tangent to the eigenvector with slope S_2 because these correspond to the case where $t \to -\infty$, as shown in Fig. 19-19*b*. Observe that in this case, the portions of the trajectories that tend to infinity are drawn parallel to the eigenvector with slope S_1 (as $t \to \infty$). Since this behavior is simply the opposite of case 1, the origin in this case is called an *unstable node.*

Case 3: Saddle point ($\lambda_1 > 0$, $\lambda_2 < 0$) For comparison purposes, let us choose the same eigenvectors shown in Fig. 19-18*c*. In this case, all trajectories must follow the direction indicated by the eigenvectors as shown in Fig. 19-19*c*. Since the trajectories have the appearance of a saddle, the origin in this case is called a *saddle point.*

19-6-2 LOCAL BEHAVIOR OF TRAJECTORIES IN THE SPIRAL CASE

An examination of the form of solutions of $x_1(t)$ and $x_2(t)$ in the right column of Table 19-1, as well as of the typical trajectories shown in Fig. 19-17*c* and *d*, shows that the trajectories in this case will never approach the singular point with a unique slope. If we substitute the expressions for $x_1(t)$ and $x_2(t)$ into Eqs. (19-85) and (19-86), we obtain the translated equations

$$y_1(t) = \rho_1 e^{\tau t} \cos(\omega t + \varphi_1) \tag{19-110}$$

$$y_2(t) = \rho_2 e^{\tau t} \cos(\omega t + \varphi_2) \tag{19-111}$$

where τ and ω are equal, respectively, to the real and imaginary parts of the natural modes λ_1 and λ_2. Again, we shall show that these natural modes completely characterize the stability of the singular point, and hence the behavior of the trajectories. We shall consider three exhaustive cases.

Case 1: Stable focus ($\tau < 0$) Since the cosine function is at most equal to unity, Eqs. (19-110) and (19-111) can be written as

$$\rho \equiv \sqrt{y_1^2(t) + y_2^2(t)} \leq \sqrt{\rho_1^2 + \rho_2^2}\, e^{\tau t} \tag{19-112}$$

where ρ can be interpreted as the distance from each point $[y_1(t), y_2(t)]$ to the origin. Since both ρ_1 and ρ_2 are independent of time, it follows that if $\tau < 0$, then $\rho \to 0$ as $t \to \infty$. Hence, the trajectory can be interpreted as a *shrinking spiral* as shown in Table 19-1 and Fig. 19-17*c*. The corresponding time waveforms shown in Table 19-1 consist of a *damped sinusoid of frequency* ω. The effect of τ is to decrease the amplitude of the sinusoid and is, therefore, called the *damping constant*. The imaginary part ω of the natural mode is usually called the *natural frequency* in this case. The singular point in this case is called a *stable focus* or a *stable spiral*. A network corresponding to this case is usually called an *overdamped network*.

Case 2: Vortex ($\tau = 0$) In this case the solutions $y_1(t)$ and $y_2(t)$ reduce to a pure sinusoidal waveform with natural frequency ω. The trajectories around the origin in this case consist of a family of concentric circles or ellipses. In view of this interpretation, the singular point corresponding to $t = 0$ is called a *vortex*.

Since there is no damping in this case, the network is usually said to be *lossless.* Obviously, this is only of theoretical interest because no physical network is truly lossless.

Case 3: Unstable focus ($\tau > 0$) In this case both $y_1(t)$ and $y_2(t)$ tend to infinity as $t \to \infty$, and the trajectory consists of an expanding spiral as shown in Table 19-1 and Fig. 19-17*d.* This type of singular point is called an *unstable focus.* A network corresponding to this case is usually called an *underdamped network.*

19-6-3 SUMMARY OF SINGULAR-POINT CLASSIFICATIONS

We have shown that a singular point of any piecewise-linear network can be classified according to whether it is a node (stable or unstable), a focus (stable or unstable), a saddle (always unstable), or a vortex (always stable). The criterion for this classification depends only on the value of the natural modes λ_1 and λ_2 of the network. Because of their frequent appearance, it is convenient for us to define

$$T \equiv (a_1 + b_2) \tag{19-113}$$

$$\Delta \equiv (a_1 b_2 - a_2 b_1) \tag{19-114}$$

TABLE 19-5 *Summary of singular-point classifications.*

Network equations:	Singular points:
$\dfrac{dx_1}{dt} = a_1x_1 + b_1x_2 + c_1$ $\dfrac{dx_2}{dt} = a_2x_1 + b_2x_2 + c_2$	$x_1 = \gamma_1,\ x_2 = \gamma_2$ $\gamma_1 = \left(\dfrac{b_1c_2 - b_2c_1}{\Delta}\right) \qquad \gamma_2 = \left(\dfrac{a_2c_1 - a_1c_2}{\Delta}\right)$ where $\Delta \equiv (a_1b_2 - a_2b_1)$ $T \equiv (a_1 + b_2)$

Type of singular point	Classification	Aperiodic case $T^2 - 4\Delta > 0$	Spiral case $T^2 - 4\Delta < 0$
Node	Stable	$\Delta > 0$ $T < 0$	
	Unstable	$\Delta > 0$ $T > 0$	
Focus	Stable		$T < 0$
	Unstable		$T > 0$
Saddle	Unstable	$\Delta < 0$	
Vortex	Stable		$T = 0$

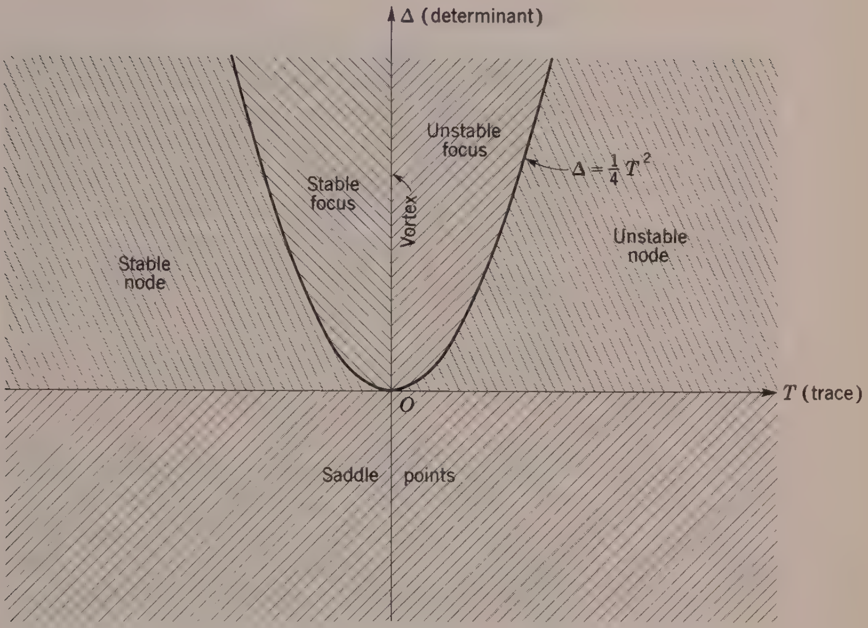


Fig. 19-20. Complete singular-point classification diagram in the T-Δ plane.

where the symbol T is sometimes called the *trace,*[1] and the symbol Δ is the *determinant* of Eqs. (19-87) and (19-88). In terms of these symbols, the complete criteria used for classifying the different types of singular points are tabulated in Table 19-5. A simple method for remembering the criteria listed in this table is to plot the parabola

$$\Delta = \tfrac{1}{4}T^2 \tag{19-115}$$

in the T-Δ plane as shown in Fig. 19-20. It is a simple matter to verify that each region shown in this figure corresponds to one and only one type of singular point. Observe that the parabola separates the nodes from the focus, the Δ axis separates the stable focus from the unstable focus, and the T axis separates the saddle points from the rest of the singular points. The points on the Δ axis in the upper half-plane correspond to vortex points. The points on the parabola as well as the points on the T axis (including the origin) correspond to $\lambda_1 = \lambda_2$, $\lambda_2 = 0$, or $\lambda_1 = \lambda_2 = 0$; all these cases have been previously excluded from consideration in Table 19-1. They represent degenerate cases and therefore require special methods to determine the nature of the singular points. Fortunately, these points are only of theoretical interest because they will never occur in any practical network, for one needs exact component values to realize degenerate singular points.

---

[1] If we write Eqs. (19-87) and (19-88) in matrix form, then the sum of the diagonal elements is usually called the "trace" in mathematics.

The above singular-point classification criteria have been derived for the case of a piecewise-linear network described by the normal-form equations (19-83) and (19-84). We shall now show that the same criteria are applicable to any normal-form equation which can be written into the form

$$\frac{dy_1}{dt} = a_1y_1 + b_1y_2 + g_1(y_1,y_2) \tag{19-116}$$

$$\frac{dy_2}{dt} = a_2y_1 + b_2y_2 + g_2(y_1,y_2) \tag{19-117}$$

where $g_1(y_1,y_2)$ and $g_2(y_1,y_2)$ represent higher-degree polynomials in y_1 and y_2. It is assumed that $g_1(0,0) = g_2(0,0) = 0$ so that the *origin* is the singular point in question. In this case, we shall classify the type of the singular point on a strictly *local basis;* that is, we shall investigate the behavior of trajectories within a very small, possibly infinitesimal, neighborhood of the origin. Under this assumption, $g_1(y_1,y_2) \approx 0$ and $g_2(y_1,y_2) \approx 0$ because y_1 and y_2 are assumed to be very small.[1] In other words, the behavior of the trajectories of Eqs. (19-116) and (19-117) in a small neighborhood of the origin is identical with the behavior of the linearized system

$$\frac{dy_1}{dt} = a_1y_1 + b_1y_2 \tag{19-118}$$

$$\frac{dy_2}{dt} = a_2y_1 + b_2y_2 \tag{19-119}$$

Hence, the classification criteria given by Table 19-5 are also applicable here. We must emphasize, however, that the trajectories will behave according to the prediction only within a small neighborhood of the origin.

Consider now the general system of normal-form equations

$$\frac{dx_1}{dt} = f_1(x_1,x_2) \tag{19-120}$$

$$\frac{dx_2}{dt} = f_2(x_1,x_2) \tag{19-121}$$

---

[1] This follows from the observation that a small number raised to a power greater than or equal to 2 becomes smaller.

The singular points are the solutions of $f_1(x_1,x_2) = 0$ and $f_2(x_1,x_2) = 0$. It suffices to consider one singular point (γ_1,γ_2) at a time. Suppose we apply the Taylor series expansion for a function of two variables about the singular point $x_1 = \gamma_1$ and $x_2 = \gamma_2$.

$$\begin{aligned} f_1(x_1,x_2) &= f_1(\gamma_1,\gamma_2) + a_1(x_1 - \gamma_1) + b_1(x_2 - \gamma_2) \\ &\quad + \text{higher-degree terms in } (x_1 - \gamma_1) \text{ and } (x_2 - \gamma_2) \end{aligned} \tag{19-122}$$

$$\begin{aligned} f_2(x_1,x_2) &= f_2(\gamma_1,\gamma_2) + a_2(x_1 - \gamma_1) + b_2(x_2 - \gamma_2) \\ &\quad + \text{higher-degree terms in } (x_1 - \gamma_1) \text{ and } (x_2 - \gamma_2) \end{aligned} \tag{19-123}$$

where

$$\begin{aligned} a_1 &= \left.\frac{\partial f_1(x_1,x_2)}{\partial x_1}\right|_{\substack{x_1=\gamma_1 \\ x_2=\gamma_2}} & b_1 &= \left.\frac{\partial f_1(x_1,x_2)}{\partial x_2}\right|_{\substack{x_1=\gamma_1 \\ x_2=\gamma_2}} \\ a_2 &= \left.\frac{\partial f_2(x_1,x_2)}{\partial x_1}\right|_{\substack{x_1=\gamma_1 \\ x_2=\gamma_2}} & b_2 &= \left.\frac{\partial f_2(x_1,x_2)}{\partial x_2}\right|_{\substack{x_1=\gamma_1 \\ x_2=\gamma_2}} \end{aligned} \tag{19-124}$$

Observe that since (γ_1,γ_2) is a singular point, $f_1(\gamma_1,\gamma_2)=f_2(\gamma_1,\gamma_2)=0$. Hence, if we translate the origin to (γ_1,γ_2) as before by defining $(x_1 - \gamma_1) = y_1$ and $(x_2 - \gamma_2) = y_2$, then Eqs. (19-120) and (19-121) become

$$\frac{dy_1}{dt} = a_1y_1 + b_1y_2 + \text{higher-degree terms in } y_1 \text{ and } y_2 \tag{19-125}$$

$$\frac{dy_2}{dt} = a_2y_1 + b_2y_2 + \text{higher-degree terms in } y_1 \text{ and } y_2 \tag{19-126}$$

Observe that Eqs. (19-125) and (19-126) are exactly of the same form as Eqs. (19-116) and (19-117). Hence, the singular-point classification criteria given in Table 19-5 are also applicable to this general case. The pertinent coefficients a_1, b_1, a_2, and b_2 are simply obtained by finding the appropriate partial derivatives at the singular point in question. The same procedure can obviously be applied to all singular points of Eqs. (19-120) and (19-121).

19-6-4 THE PHASE PORTRAIT

If the origin is the only singular point, as is true in the case of a second-order linear network without independent sources, then the system of trajectories corresponding to each type of singular points as shown in Fig. 19-19 will provide a complete "picture" of the network's behavior. If the linear network contains independent sources, there will still be only one singular point located at $x_1 = \gamma_1$ and $x_2 = \gamma_2$. In this case, the system of trajectories in Fig. 19-19 can be simply translated from the origin to the singular point (γ_1,γ_2). We shall refer to the system of trajectories which depict the network's behavior in the entire phase plane as the *phase portrait*.

If the network is piecewise-linear or nonlinear, there will generally be more than one singular point. In this case the phase portrait becomes more complicated, and more interesting, too; but in any case it can be obtained by applying the properties developed in the previous sections. The systematic procedures for obtaining the phase portrait associated with a nonlinear network can be outlined as follows:[1]

Step 1. Find the location of all singular points.

Step 2. Classify the type of each singular point by the criteria given in Table 19-5.

Step 3. If a singular point (γ_1,γ_2) is either a node or a saddle, determine the limiting slopes S_1 and S_2 and plot the corresponding eigenvectors with the proper direction of motion indicated by an arrow.

Step 4. Apply the trajectory slope theorem of Sec. 19-6-1 to sketch the system of trajectories around a small neighborhood of each node and each saddle.

Step 5. Sketch the system of spiral trajectories around a small neighborhood of each focus.

Step 6. Based on the systems of trajectories drawn around the neighborhood of all singular points, it is usually possible to draw a system of "connecting" trajectories that will go from one singular point to another. The resulting phase portrait, of course, gives only an approximate picture, but this is often satisfactory for most practical purposes. After all, the main value of the phase portrait is to provide qualitative, not quantitative, information concerning the network's behavior over the entire phase plane.

Exercise 1: Derive Eqs. (19-87) and (19-88) by substituting Eqs. (19-85) and (19-86) into Eqs. (19-83) and (19-84).

Exercise 2: Derive Eqs. (19-94) and (19-95) and verify that the slopes of approach are equal numerically to the natural modes for any network described by $dx_1/dt = x_2$, $dx_2/dt = f(x_1,x_2)$.

---

[1] A user-oriented computer program for carrying out a phase-plane analysis has been developed for this purpose. For details of this program, see L. O. Chua and F. B. Moyer, Computer-Aided Phase Plane Analysis, *12th Midwest Symp. Circuit Theory, University of Texas, Austin, Tex.*, pp. III.2.1–III.2.12, April 21–22, 1969.

Exercise 3: By an appropriate choice of parameter values in Eqs. (19-104) and (19-105), derive the following useful relationships:

(*a*) $$\lim_{t\to\infty} \frac{y_2(t)}{y_1(t)} = \frac{\alpha_2}{\alpha_1} = \frac{a_2}{\lambda_1 - b_2} = S_1$$

(*b*) $$\lim_{t\to-\infty} \frac{y_2(t)}{y_1(t)} = \frac{\beta_2}{\beta_1} = \frac{a_2}{\lambda_2 - b_2} = S_2$$

Exercise 4: Prove that only four trajectories can pass through a saddle point, namely, the eigenvectors.

Exercise 5: Sketch the general form of the trajectories around the three types of singular points of the equations $dx_1/dt = x_2$ and $dx_2/dt = f(x_1,x_2)$. Verify that the eigenvectors corresponding to a stable node must always lie in the second and the fourth quadrants.

Exercise 6: Verify the relationships given in Eq. (19-106) and (19-107) by using the data obtained in the example given in Sec. 19-5.

Exercise 7: Verify the entries in Table 19-5 and the geometrical interpretation given in the T-Δ plane shown in Fig. 19-20.

Exercise 8: Sketch the phase portrait for the network shown in Fig. 19-14*a*.

19-7 SUMMARY

Basic philosophy An autonomous second-order nonlinear network is best analyzed in two phases. The first phase consists of a qualitative analysis wherein the general behavior of the network over all possible initial states is determined. In this phase of analysis a certain amount of inaccuracies can be tolerated because we are looking at the network from a global or macroscopic point of view. Once the general properties of the trajectories are known, we can then proceed to the second phase. In this phase of analysis we seek to find the exact solution waveforms corresponding to some meaningful choice of initial states. The initial state to be chosen depends on what we are looking for. If, from the qualitative analysis, we found the network had a stable limit cycle, then we might be interested in finding the steady-state solution by choosing an initial state which falls on the limit cycle. On the other hand, if we are interested primarily in the transient behavior of a network during a switching operation, then the initial state ought to be chosen accordingly.

Qualitative analysis This is obtained by constructing the vector field associated with the normal-form equations of the network. The vector field can be obtained by either a brute-force method requiring the use of a computer with plotting capability or by the isocline method. In order to be practical, the isoclines of a nonlinear network can be obtained by applying the *isocline network theorem.* This theorem establishes the important relationship between the isocline of a dynamic network and the TC plots of the associated resistive isocline network.

Another, less accurate but simpler, approach for obtaining the qualitative properties of the solution is to construct the phase portrait associated with the network. The phase portrait is simply a collection of trajectories corresponding to different initial states, and can therefore be constructed from the vector field. However, a simpler approximate procedure would be to find the location and type of all singular points. The system of trajectories can be sketched by applying the *trajectory slope theorem* and the other properties developed in this chapter.

Quantitative analysis If the normal-form equations can be obtained without much difficulty, then the numerical techniques can be used for obtaining the solution waveforms corresponding to any initial state. A more practical approach would be to apply the piecewise-linear method developed in Sec. 1. The key to this method is the construction of the switching plane associated with the network. One advantage of this method over numerical techniques is that the solutions can be obtained in a semianalytical form, thereby providing much more insight into the effect of some circuit parameters on the networks' behavior.

PROBLEMS

19-1 A simple "compass and ruler" technique called *Lienard's method* can be used to construct the trajectories, and hence the phase portrait, associated with any network which is described by the normal-form equation $dx_1/dt = x_2$ and $dx_2/dt = -f(x_2) - x_1$. The method is based on the observation that the slope

$$\frac{dx_2}{dx_1} = -\frac{[x_1 + f(x_2)]}{x_2} = m$$

at each point $P(x_1,x_2)$ in the phase plane can be obtained by plotting first the curve $x_1 = -f(x_2)$ and then drawing a director perpendicular to the line connecting points P and Q, as shown in Fig. P19-1.

(a) Based on the graphical construction shown in Fig. P19-1, describe the procedure for drawing the trajectory starting at any initial state P_0 by using only a compass and a ruler.

(b) Construct the phase portrait by the above procedure for the linear system $dx_1/dt = x_2$ and $dx_2/dt = -x_1$.

(c) Use Lienard's method to obtain the limit cycle of Rayleigh's equation $dx_1/dt = x_2$ and $dx_2/dt = \epsilon(x_2 - \frac{1}{3}x_2^3) - x_1$ with $\epsilon = 0.3$.

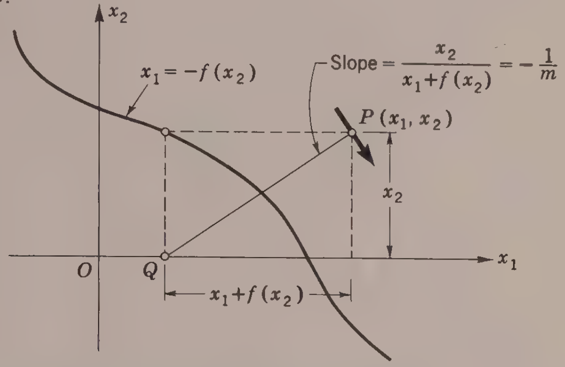


Fig. P19-1.

19-2 Show that Lienard's method described in Prob. 19-1 can be generalized so that it can be applied to the normal-form equations $dx_1/dt = x_2$ and $dx_2/dt = -f(x_2) - g(x_1)$ by plotting the *two* curves $x_1 = -f(x_2)$ and $x_2 = -g(x_1)$ in the phase plane.

19-3 Repeat Prob. 19-2 for the normal-form equations $dx_1/dt = x_2 + g(x_1)$ and $dx_2/dt = -x_1 - f(x_2)$.

19-4 Repeat Prob. 19-2 for the normal-form equations $dx_1/dt = x_2 + g(x_1)$ and $dx_2/dt = -x_1 - f(x_1)$.

19-5 A limit cycle Γ is said to be stable (unstable) if all trajectories starting inside and outside of Γ converge to (diverge from) Γ as $t \to \infty$. It is said to be semistable if all trajectories starting inside of Γ converge (diverge) and those starting outside of Γ diverge (converge) from Γ, as $t \to \infty$.

(a) Sketch a typical phase portrait associated with each of these three types of limit cycles.

(b) It often turns out that a nonlinear system possesses several concentric or "nested" limit cycles. Sketch a typical phase portrait with this property and verify that the limit cycles in this case must alternate from stability to instability.

(c) As the system parameters are varied continuously, the limit cycles will move from their original positions. Use this observation to demonstrate how a semistable limit cycle may occur at some critical parameter values.

19-6 If the normal-form equations $dx_1/dt = f_1(x_1,x_2)$ and $dx_2/dt = f_2(x_1,x_2)$ possess the property that the expression

$$F(x_1,x_2) = \frac{\partial f_1(x_1,x_2)}{\partial x_1} + \frac{\partial f_2(x_1,x_2)}{\partial x_2}$$

never changes sign in a region (without holes) D of the phase plane, then no limit cycle can exist in D. This is known as *Bendixon's theorem.*

(a) Prove Bendixon's theorem by applying Green's theorem given in most advanced calculus or electromagnetic field theory books.

(b) Apply Bendixon's theorem to prove that the Van der Pol equation cannot have a limit cycle within the region D defined by $-1 < x_1 < 1$, $-\infty < x_2 < \infty$.

(c) Use Bendixon's theorem to prove that a capacitor-loaded second-order network cannot possess a limit cycle if the incremental conductances di_1/dv_1 and di_2/dv_2 of the dc-resistive nonlinear network N in Fig. 19-5*a* (with port 2 and port 1 terminated, respectively, by a battery of voltage E) are positive for all values of E.

(d) Give the results corresponding to (*c*) for the inductor-loaded case.

(e) Repeat (*c*) for the inductor-capacitor-loaded case.

19-7 Prove that Bendixon's theorem remains valid if $F(x_1,x_2)$ in Prob. 19-6 is replaced by

$$F(x_1,x_2) = \frac{\partial[g(x_1,x_2)f_1(x_1,x_2)]}{\partial x_1} + \frac{\partial[g(x_1,x_2)f_2(x_1,x_2)]}{\partial x_2}$$

where $g(x_1,x_2)$ is any continuously differentiable functions of x_1 and x_2 in the region D.

19-8 Prove that if the two capacitors of an autonomous second-order network are characterized by charge-controlled v-q curves, then the isoclines in the q_1-q_2 phase plane can be determined in three steps. (*a*) Solve for the v_2-vs.-v_1 TC plots of the normalized isocline network 1 shown in Fig. P19-8*a* or the v_1-vs.-v_2 TC plot of the normalized isocline network 2 shown in Fig. P19-8*b*. (*b*) Perform a *graphical composition* between the v_2-q_2 curve of capacitor C_2 with each v_2-vs.-v_1 TC plot or a *graphical elimination* between the v_2-q_2 curve with each v_1-vs.-v_2 TC plot. The result is a curve in the q_2-vs.-v_1 plane. (*c*) Perform a *graphical elimination* between each q_2-vs.-v_1 curve obtained from step 2 with the v_1-q_1 curve of capacitor C_1. The result is the desired isocline in the q_1-q_2 plane.

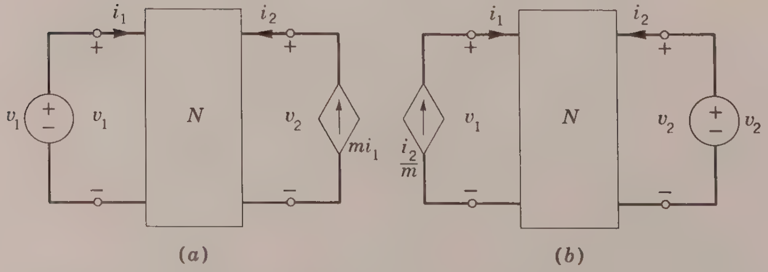


Fig. P19-8.

19-9 By a procedure similar to that given in Prob. 19-8, derive a method for constructing the isoclines associated with a flux-controlled inductor-loaded network from the normalized isocline network 1 shown in Fig. P19-9a or the normalized isocline network 2 shown in Fig. P19-9b.

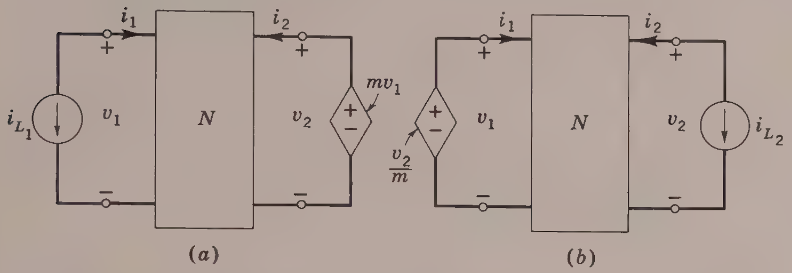


Fig. P19-9.

19-10 By a procedure similar to that given in Prob. 19-8, derive a method for constructing the isoclines associated with a charge-controlled capacitor and flux-controlled inductor-loaded network from the normalized isocline network 1 shown in Fig. P19-10a or the normalized isocline network 2 shown in Fig. P19-10b.

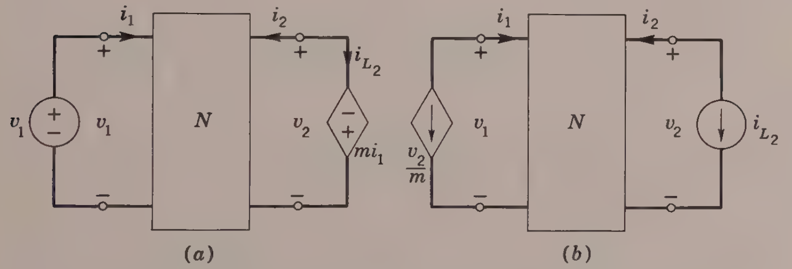


Fig. P19-10.

19-11 Prove that if the capacitor-loaded network shown in Fig. P19-11 possesses a flip-over symmetry with respect to a vertical axis through the center of N, then all isoclines in the v_1-v_2 plane are symmetrical with respect to the 45° straight line through the origin. Moreover, the slopes m and m' of each symmetric pair of isoclines are related by $m' = 1/m$. This result is referred to as the *vertically symmetric isocline network theorem.* HINT: Consider *both* isocline networks (but label the variables of one of them by primes) and then flip over the unprimed isocline network about the axis of symmetry.

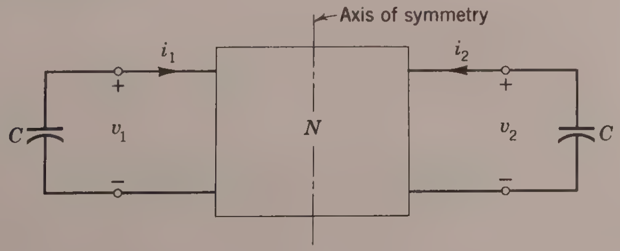


Fig. P19-11.

19-12 Prove that if the capacitor-loaded network shown in Fig. P19-11 possesses rotational symmetry with respect to an axis perpendicular to the center of N, then all isoclines in the v_1-v_2 plane are symmetrical with respect to the 135° straight line through the origin. Moreover, the slopes m and m' of each symmetric pair of isoclines are related by $m' = -1/m$. This result is referred to as the *rotationally symmetric isocline network theorem.*

19-13 Derive the *vertically symmetric isocline network theorem* (see Prob. 19-11) and the *rotationally symmetric isocline network theorem* (see Prob. 19-12) for the following:

(a) The inductor-loaded network.
(b) The inductor-capacitor-loaded network.

19-14 Classify the singular points of the tunnel-diode network shown in Fig. 13-18*c* of Chap. 13. Sketch the phase portrait of this network by the technique described in Sec. 19-6-4.

19-15 Classify the singular points of the twin tunnel-diode switching network shown in Fig. 19-11*a*. Sketch the phase portrait of this network by the technique described in Sec. 19-6-4.

19-16 Derive the switching plane associated with the twin tunnel-diode switching network shown in Fig. 19-11*a*. Find the solution waveforms corresponding to the initial state P_c shown in Fig. 19-12.

19-17 The vacuum-tube circuit shown in Fig. P19-17 is called the *Eccles-Jordan* or *cross-coupled multivibrator* and is the workhorse of many instrumentation circuits.

(a) Find the singular points of this circuit.
(b) Calculate the coefficients a_1, b_1, a_2, and b_2 for Eq. (19-89) corresponding to each singular point of this circuit.
(c) Classify each singular point with the help of Table 19-5.

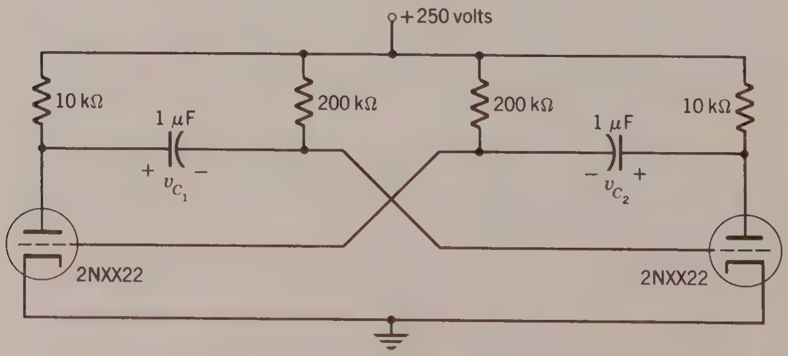


Fig. P19-17.

19-18 The circuit shown in Fig. P19-18 is the transistorized version of the Eccles-Jordan multivibrator circuit. (See Prob. 19-17.)

(a) Find the singular points of this circuit.
(b) Calculate the coefficients a_1, b_1, a_2, and b_2 for Eq. (19-89) corresponding to each singular point of this circuit.
(c) Classify each singular point with the help of Table 19-5.

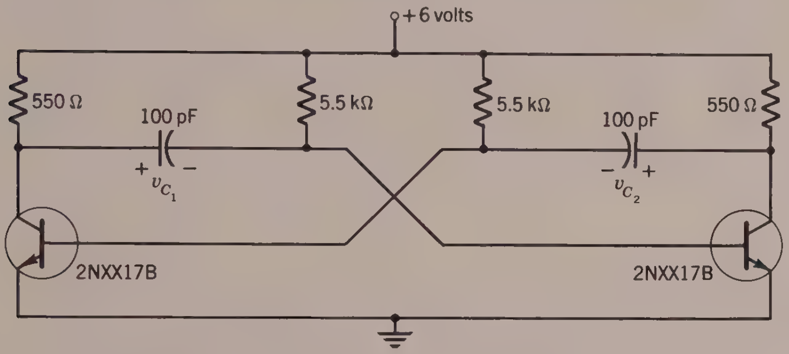


Fig. P19-18.

19-19 (a) Show that a *nonautonomous first-order system* described by $dx/dt = f(x,t)$ can be analyzed by an equivalent *autonomous second-order system* described by

$$\frac{dx}{dt} = f(x,y) \qquad \frac{dy}{dt} = 1$$

(b) Find the singular points of the equivalent second-order system.

19-20 A computer program for constructing phase portraits can be used to analyze a nonautonomous first-order network N by transforming it into an equivalent autonomous second-order network N^*.
(a) Show that N^* can be constructed from N by adding only one energy-storage element and one independent current source. HINT: Use the result of Prob. 19-19 to transform each ac-independent source in N into an appropriate nonlinear controlled source.
(b) In order to apply the iterative piecewise-linear method of analysis, show that each *nonlinear controlled source* can be modeled by an appropriate nonlinear resistor and two *linear controlled sources.*

20 ANALYSIS OF *n*TH-ORDER NONLINEAR NETWORKS

20-1 BEYOND THE SECOND-ORDER AUTONOMOUS NETWORK BARRIER

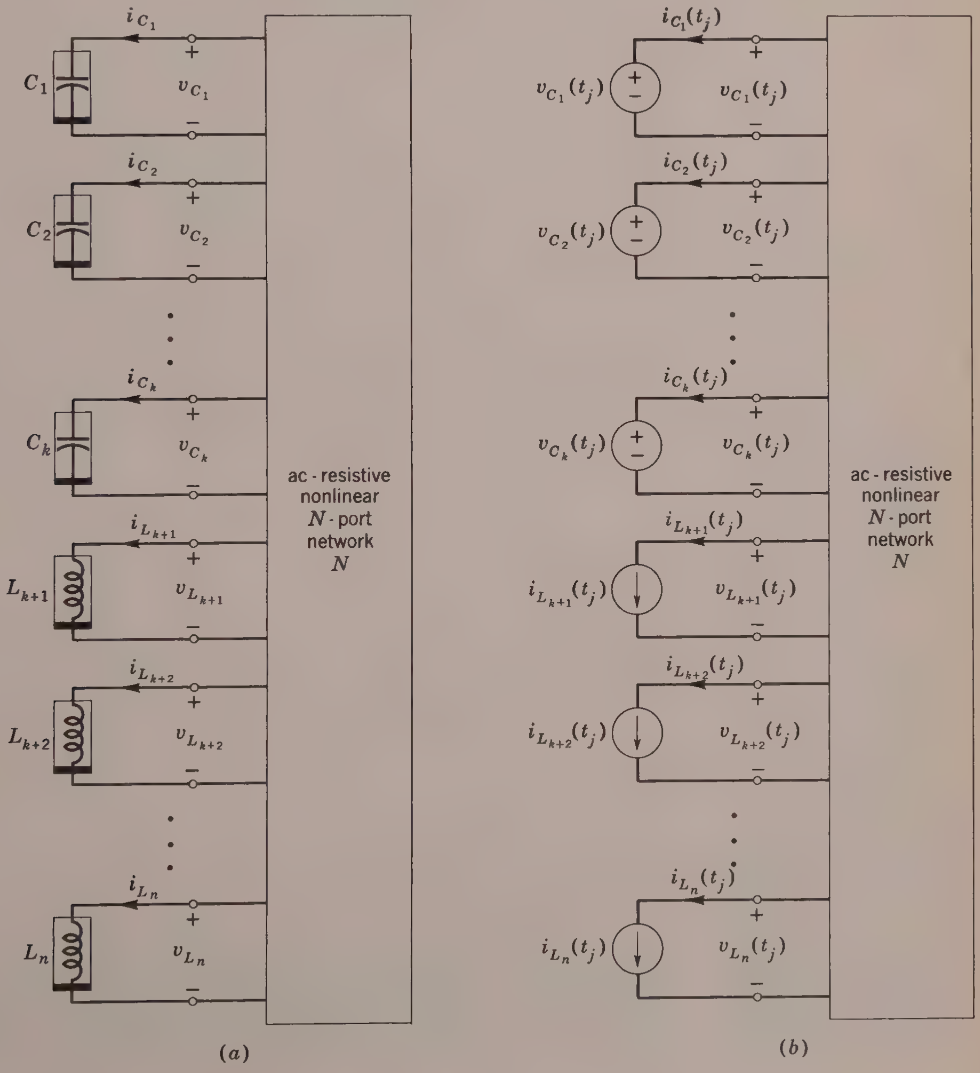
So far we have been able to develop several methods for analyzing first-order autonomous and nonautonomous networks, as well as second-order autonomous networks. Most of the qualitative properties of these networks have been derived from simple geometrical interpretations. Unfortunately, beyond this class of dynamic networks lies the huge class of higher-order networks for which geometrical methods and interpretations are less useful, if not completely useless. This is due to the necessity to consider *higher-dimensional spaces,* thereby losing the simplicity inherent in the two-dimensional plane. For example, the concept of drawing the phase portrait associated with each type of singular point is no longer applicable in higher-dimensional networks. Neither does the exhaustive classification of singular points into a node, focus, saddle, or vortex admit to an *n*-dimensional generalization. In view of these difficulties, the analysis of any higher-order dynamic nonlinear network invariably requires the use of a digital computer.

It is clearly unrealistic to attempt to cover the complex subject of higher-order network analysis in one chapter. Our object in this final chapter will, therefore, be a modest one. In contrast with the detailed exposition given in the preceding chapters, we shall merely present a survey of various techniques for analyzing practical *n*th-order nonlinear networks. These techniques can be categorized into five different approaches: (1) the numerical approach, (2) the piecewise-linear approach, (3) the heuristic approach, (4) the qualitative approach, and (5) the approximate analytical approach.

20-2 THE NUMERICAL-INTEGRATION APPROACH

Any *n*th-order nonlinear network can be represented as shown in Fig. 20-1*a*, where the *n* energy-storage elements have been brought out explicitly across an *n*-port ac-resistive nonlinear network *N*. The normal-form equations describing this network, if they exist, are given by

$$\frac{dx_i}{dt} = f_i(x_1,x_2, \ldots ,x_n,t) \qquad i = 1, 2, \ldots, n \tag{20-1}$$



(*a*) (*b*)

Fig. 20-1. An *n*th-order nonlinear network and its equivalent representation as a resistive *n*-port network terminated by appropriate independent voltage sources and current sources.

where the state variables x_i may denote capacitor voltages or charges for $i \leq k$ and inductor currents or flux linkages for $k < i \leq n$.

In theory the solution of Eq. (20-1) corresponding to any initial state can be found by the numerical techniques described in Sec. 4-7. In practice, however, there are two serious problems that must be overcome. The first problem is the derivation of the normal-form equation in analytic form. This is often a difficult, if not impractical, task because it entails not only the fitting of each nonlinear characteristic curve by an equation, but also the solution of a system of *nonlinear* transcendental equations. The second problem is even more fundamental in nature; namely, the normal-form equation may not exist as a result of the choice of incomplete network models (see Sec. 14-5). A discussion of the conditions for guaranteeing the existence of normal-form equations is beyond the scope of this book.[1] However, as a practical guide it is safe to say that if well-established network models, which usually contain at least one parasitic inductance and one parasitic capacitance, are chosen, then most likely the normal-form equation will exist. This leaves us with only the first problem to worry about.

Our experience with the constant-slope network theorem and the isocline network theorem suggests that we may be able to overcome this problem by avoiding it. Suppose we write the equations of motion of the network shown in Fig. 20-1*a* in terms of the state variables v_{C_j} and i_{L_j}; namely,

$$\begin{aligned}
\frac{dv_{C_1}}{dt} &= \frac{1}{C(v_{C_1})} i_{C_1} \\
&\cdots\cdots\cdots\cdots \\
\frac{dv_{C_k}}{dt} &= \frac{1}{C(v_{C_k})} i_{C_k} \\
\frac{di_{L_{k+1}}}{dt} &= \frac{1}{L(i_{L_{k+1}})} v_{L_{k+1}} \\
&\cdots\cdots\cdots\cdots \\
\frac{di_{L_n}}{dt} &= \frac{1}{L(i_{L_n})} v_{L_n}
\end{aligned} \qquad (20\text{-}2)$$

where $C(v_{C_j}) \equiv dq(v_{C_j})/dv_{C_j}$ and $L(i_{L_j}) \equiv d\varphi(i_{L_j})/di_{L_j}$ are the *incremental capacitance* and the *incremental inductance*, respectively. Suppose we are given the initial states $v_{C_1}(t_0), v_{C_2}(t_0), \ldots, v_{C_k}(t_0)$,

---

[1] A general treatment of this subject can be found in L. O. Chua and R. A. Rohrer, On the Dynamic Equations of a Class of Nonlinear *RLC* Networks, *IRE Trans. Circuit Theory*, vol. CT-12, no. 4, pp. 474–489, December, 1965.

$i_{L_{k+1}}(t_0)$, $i_{L_{k+2}}(t_0)$, . . . , $i_{L_n}(t_0)$ and are required to find the solutions during some prescribed time interval $t_0 \leq t \leq t_f$. Observe that if we replace each capacitor by a voltage source $v_{C_j}(t_0), j = 1, 2, \ldots, k$ and each inductor by a current source $i_{L_j}(t_0), j = k+1, k+2, \ldots, n$, we would obtain the resistive network shown in Fig. 20-1*b* (with $t_j = t_0$). We can, therefore, develop an automatic computer program to solve for the capacitor currents $i_{C_j}(t_0)$ and the inductor voltages $v_{L_j}(t_0)$ at t_0. If we divide the time interval (t_0, t_f) into discrete increments of time with a step size $\Delta t = h$, then the solution of Eq. (20-2) at $t_1 = t_0 + h$ can be approximately determined by applying the Euler algorithm[1] described in Eq. (4-61) or Sec. 4-7-1; namely,

$$
\begin{aligned}
v_{C_1}(t_1) &= v_{C_1}(t_0) + h\left[\frac{1}{C(v_{C_1}(t_0))}\, i_{C_1}(t_0)\right] \\
v_{C_2}(t_1) &= v_{C_2}(t_0) + h\left[\frac{1}{C(v_{C_2}(t_0))}\, i_{C_2}(t_0)\right] \\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
v_{C_k}(t_1) &= v_{C_k}(t_0) + h\left[\frac{1}{C(v_{C_k}(t_0))}\, i_{C_k}(t_0)\right] \\
i_{L_{k+1}}(t_1) &= i_{L_{k+1}}(t_0) + h\left[\frac{1}{L(i_{L_{k+1}}(t_0))}\, v_{L_{k+1}}(t_0)\right] \\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
i_{L_n}(t_1) &= i_{L_n}(t_0) + h\left[\frac{1}{L(i_{L_n}(t_0))}\, v_{L_n}(t_0)\right]
\end{aligned}
\tag{20-3}
$$

The above procedure can, of course, be iterated to find the values of v_{C_j} and i_{L_j} at $t = t_2, t_3, \ldots, t_f$. Hence, to find the values of $v_{C_1}(t_{j+1}), \ldots, v_{C_k}(t_{j+1}), i_{L_{k+1}}(t_{j+1}), \ldots, i_{L_n}(t_{j+1})$, we simply solve for the capacitor currents $i_{C_1}(t_j), \ldots, i_{C_k}(t_j)$ and the inductor voltages $v_{L_{k+1}}(t_j), \ldots, v_{L_n}(t_j)$ of the *resistive* network shown in Fig. 20-1*b* at $t = t_j$. If we assume a uniform step size h, then a total of

$$N = \frac{(t_f - t_0)}{h}$$

iterations will be required to obtain the solutions over the prescribed time interval. Obviously, in order to be practical, the above procedures must be implemented automatically on a computer. Several such user-oriented programs are currently available, and no doubt many faster and more general programs will become widely available in the near future.[2]

---

[1] In view of its simplicity, the Euler algorithm is chosen here mainly for pedagogical reasons. In practice, more accurate algorithms are usually used in automatic network analysis computer programs.

[2] These user-oriented programs are all written under the assumptions that the normal-form equation exists and that the network possesses a unique solution. This is, of course, not true in general. For a more detailed discussion of the subtle points involved in the development of any general user-oriented computer programs, see Franklin F. Kuo and Waldo G. Magnuson, Jr., "Computer Oriented Circuit Design," Prentice-Hall Inc., Englewood Cliffs, N.J., 1969.

20-3 THE PIECEWISE-LINEAR APPROACH

The numerical-integration approach suffers from at least two very serious problems. The first problem concerns the accuracy of the solution. Since the integration formula is never exact, the solution obtained by any numerical technique is always subject to numerical inaccuracy, and the basic problem is to keep the errors from building up. One obvious method is to keep the step size down to as small a value as possible. But this also increases the time it takes to complete the iterations. This leads to the second problem, namely, to find more efficient methods for solving the resistive network in Fig. 20-1*b*. Remember, *this network must be solved during each iteration* and is therefore the main time killer in the entire algorithm.

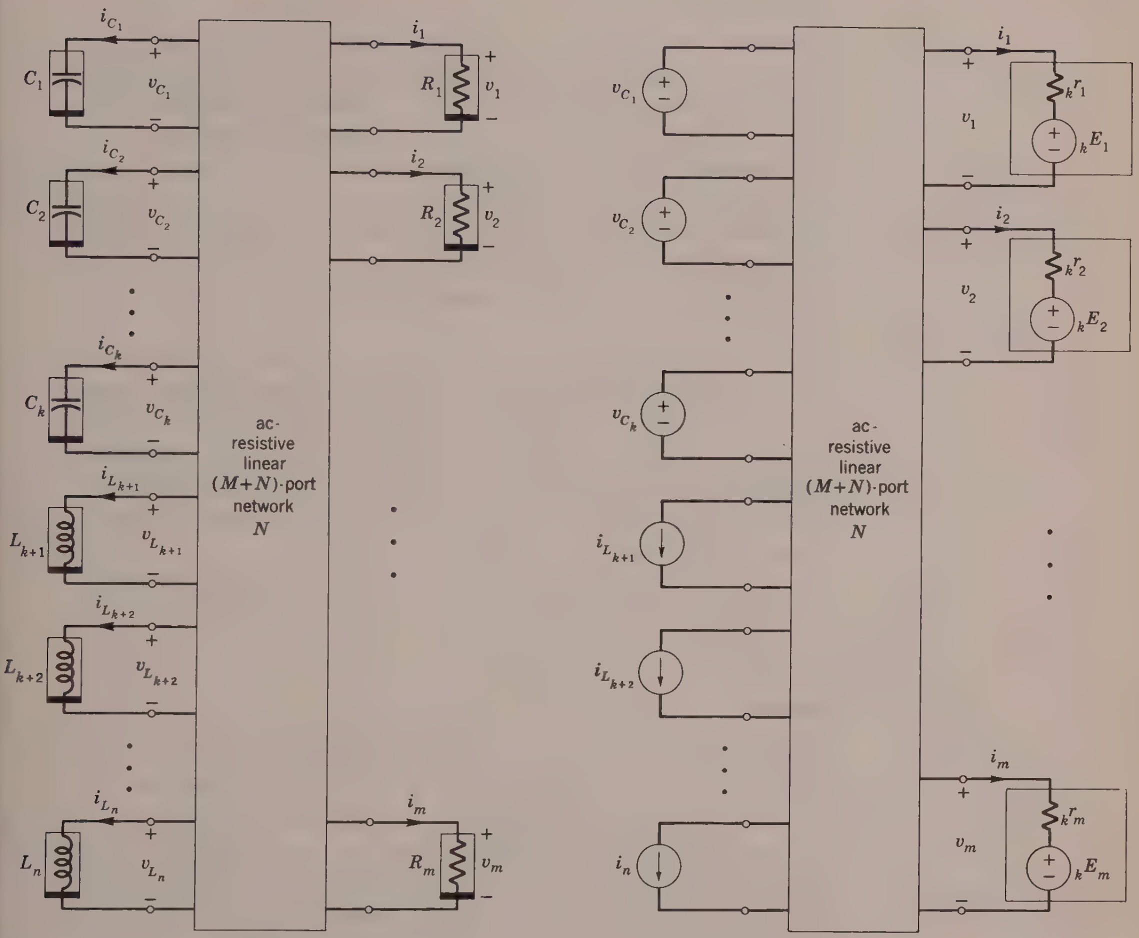
An alternate method which is sometimes more practical is the piecewise-linear approach. Suppose we extract all nonlinear resistors from the network N of Fig. 20-1*a* as shown in Fig. 20-2*a*. From this we obtain the iterative piecewise-linear network shown in Fig. 20-2*b*. Notice that the $(m + n)$-port network N is a linear resistive network, and hence it is a simple matter to derive the transfer voltage or current relationships from the nonlinear resistors in terms of the capacitor voltages v_{C_j} and the inductor currents i_{L_j}. If the network is *autonomous,* then these relationships assume the general form

$$\begin{aligned}
v_1 &= H_{11}v_{C_1} + H_{12}v_{C_2} + \cdots + H_{1k}v_{C_k} + H_{1(k+1)}i_{L_{k+1}} + \cdots + H_{1n}i_{L_n} + K_1 \\
v_2 &= H_{21}v_{C_1} + H_{22}v_{C_2} + \cdots + H_{2k}v_{C_k} + H_{2(k+1)}i_{L_{k+1}} + \cdots + H_{2n}i_{L_n} + K_2 \\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
v_m &= H_{m1}v_{C_1} + H_{m2}v_{C_2} + \cdots + H_{mk}v_{C_k} + H_{m(k+1)}i_{L_{k+1}} + \cdots + H_{mn}i_{L_k} + K_m
\end{aligned} \tag{20-4}$$

where H_{ij} and K_i are constants. By imposing the constraint

$$_kE_j^- < v_j < {}_kE_j^+ \tag{20-5}$$

corresponding to each pertinent segment k of resistor R_j, we obtain a set of inequalities defining a *cell* in the n-dimensional space in exact analogy with the two-dimensional case. Since the network is linear while operating in a cell, the solution can be obtained by a number of more accurate and less time-consuming numerical methods. The concept of a switching space is clearly applicable here, only this time we must use a computer to check *when* the boundary of a particular cell is intersected by the trajectory.



(*a*) (*b*)

Fig. 20-2. An *n*th-order nonlinear network and its associated iterative piecewise-linear network.

The complete procedure for obtaining the solution is similar to the two-dimensional case described in Sec. 19-3.

If the network is nonautonomous, the procedure remains valid, only this time the coefficients H_{ij} and K_i will be functions of time. But since a computer is being used anyway, these time-varying coefficients do not create any new problem.[1]

20-4 THE HEURISTIC APPROACH

In many practical networks it is possible to arrive at a satisfactory analysis by making use of some simplifying assumptions or idealizations. If a network is not too complex, it is sometimes possible to predict its behavior by appealing to the physics of the devices. On other occasions a network may possess certain special features, such as symmetry, which will lead to a drastic simplification in the analysis. The ability to uncover these special features and to take advantage of them depends a great deal on the engineer's experience and intuition. Since most of these techniques are of a very specialized, if not approximate, nature, it is not possible at present to unify them into a well-organized theory. In the absence of a better name, these techniques will be referred to as the *heuristic approach.*

---

[1] An alternate procedure in the nonautonomous case consists of extracting all ac sources in a manner similar to that described in Sec. 18-3-2. If there are α ac sources, then Eq. (20-4) will be a function of α additional variables. However, the coefficients H_{ij} and K_i will now be constants.

It must be emphasized that in many practical networks the heuristic approach is not only easier but also more illuminating. We should, therefore, never refuse to learn a particular method only because it is not of sufficient generality. On the contrary, these methods are usually extremely valuable, provided the user knows exactly the simplifying assumptions and limitations that invariably characterize them.

In order to illustrate the value of the heuristic approach, let us consider an example.

EXAMPLE

Consider the tunnel-diode multivibrator circuit shown in Fig. 16-10*a*, which we reproduce as Fig. 20-3*a* for convenience. The dynamic route of this network exhibits horizontal jumps as shown in Fig. 16-10*b*. This corresponds to instantaneous jumps in the inductor voltage waveform $v_L(t)$ as shown in Fig. 16-11*b*. Earlier, in Sec. 14-6, we postulated the jump phenomenon in order to bypass the multivalued character of the normal-form equation. We have also shown that a more satisfactory, though more complicated, approach is to introduce one or more parasitic elements so that the normal-form equation becomes single-valued. For example, if we introduce a parasitic capacitor across the tunnel diode as shown in Fig. 20-3*b*, the normal-form equation is easily found to be given by

$$\frac{di_L}{dt} = \frac{v_C - Ri_L - [E + v_s(t)]}{L} \tag{20-6}$$

$$\frac{dv_C}{dt} = -\frac{[g(v_C) + i_L]}{C} \tag{20-7}$$

where $g(v_C)$ denotes the tunnel diode v_d-i_d curve. Equations (20-6) and (20-7) represent a system of nonautonomous normal-form

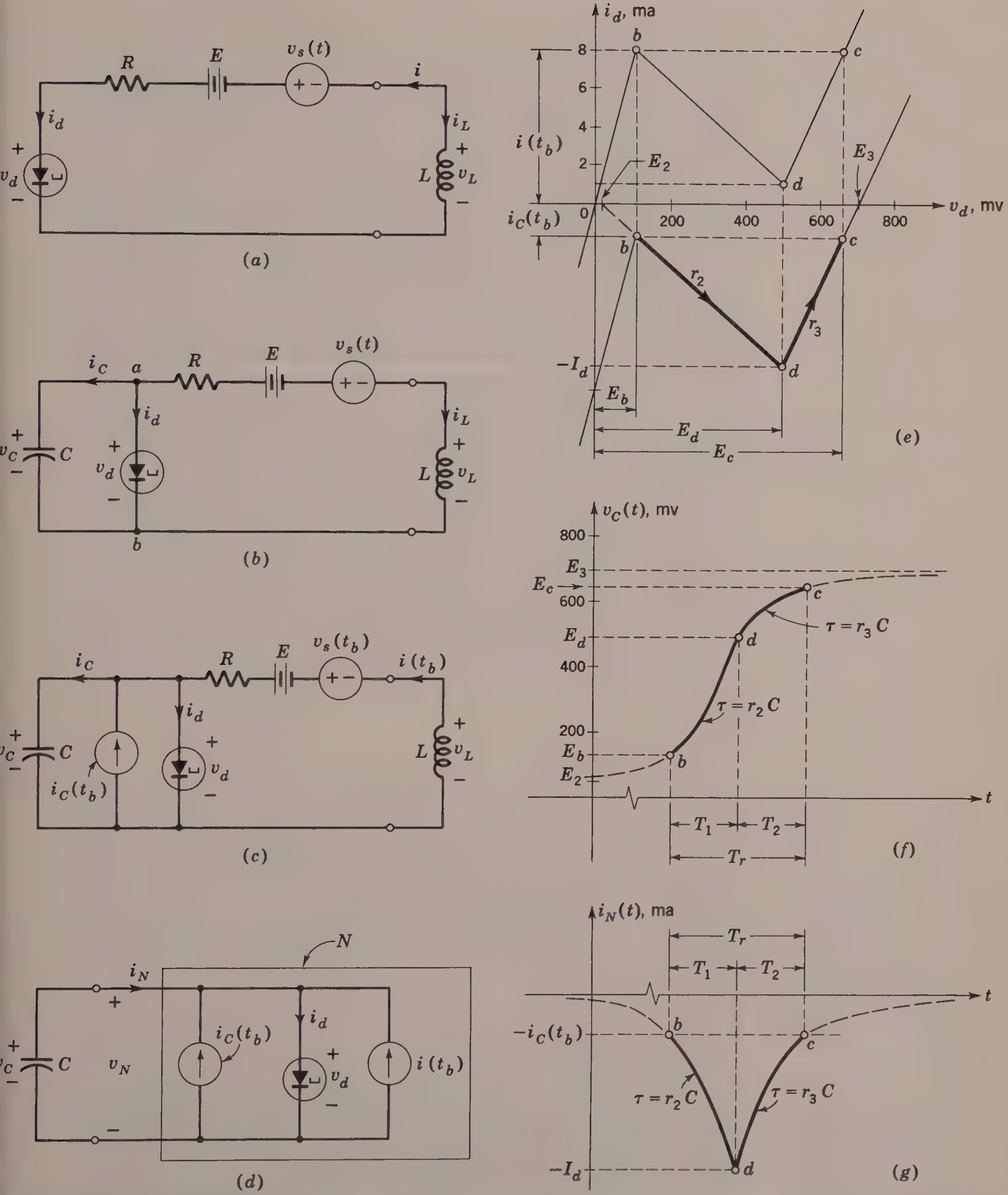


Fig. 20-3. The step-by-step procedure for analyzing a second-order nonautonomous nonlinear network as two first-order autonomous networks.

equations, and can be solved by either the numerical-integration approach or the piecewise-linear approach. The resulting solution will be very nearly that obtained in Fig. 16-11. The only significant difference lies in the fact that instead of a vertical jump, the solution waveform changes extremely fast.

Heuristically, it seems that the parasitic capacitor plays a significant role only during the fast-changing phase of the solution waveform. This is a reasonable assumption because the value of C being very small (a few picofarads), the current $i_C = C\,dv_C/dt$ will also be negligibly small, unless dv_C/dt is very large. Hence, we conclude that we may replace the capacitor by an open circuit most of the time, and the dynamic route shown in Fig. 16-10*b* is, therefore, also applicable most of the time except during jumps. Observe that if we assume that the dynamic route jumps instantaneously at $t = t_b$, then $dv_L(t)/dt = \infty$, and since $v_C = Ri_L + E + v_s(t) + v_L$, it follows that $dv_C/dt = \infty$ at $t = t_b$. This means that even though C is very small, $i_C = C\,dv_C/dt$ is *not negligible* at $t = t_b$. Hence, to calculate the waveform during the transition from point b to point c as shown in Fig. 16-10*b*, we must put the capacitor back. Observe, however, that since the solution waveforms in Fig. 16-11 are obtained with the *assumption* that $i_C = 0$, if we connect the capacitor suddenly across the tunnel diode at $t = t_b$ as shown in Fig. 20-3*b*, the currents at node a will *not* satisfy KCL because

$$i_C(t_b) = C\frac{dv_C}{dt}\bigg|_{t=t_b} \neq 0 \tag{20-8}$$

In order to remove this inconsistency, we may connect a *compensating current source*[1] with a terminal current equal to $i_C(t_b)$ as shown in Fig. 20-3*c*. We can, therefore, consider this as an "approximate" equivalent circuit that applies during the very short transition time T_r it takes the dynamic route in Fig. 16-10*b* to move from point b to point c. From experience, as well as from intuition, we know

$$T_r \ll \tau_1 \tag{20-9}$$

where τ_1 is the time constant of the inductor.[2] This means that for most practical purposes, the inductor current will change very slightly during the transition time interval T_r, and hence for purposes of analysis we may replace the inductor by a current source with a terminal current equal to $i(t_b)$ during the entire interval T_r, as shown in Fig. 20-3*d*. Observe that the resistor-battery-triggering-source combination in the original circuit is redundant during this time and has been replaced by a short circuit.[1]

---

[1] Observe that this is a fictitious current source which we introduced in order to avoid violating KCL. After all, even an approximate method must be consistent with both KCL and KVL at all times.

[2] Observe that the time scale in the waveform shown in Fig. 16-11 must be greatly expanded in the vicinity of t_b in order to be able to mark off the corresponding transition time T_r.

The circuit shown in Fig. 20-3*d* is an approximate equivalent circuit that applies *only* during the transition time T_r. Observe that it is a first-order autonomous network. Hence, we can apply the techniques developed in Chap. 14 to obtain the capacitor voltage and current waveforms. First, we must obtain the DP plot across the black box N in Fig. 20-3*d*. This is shown in Fig. 20-3*e*. The dynamic route starting at the initial state b is readily sketched as shown in this figure. Recall that the circuit shown in Fig. 20-3*d* is valid only up to the time when the dynamic route arrives at point c. After this time the transition is completed, and dv_C/dt is no longer significant. This means that the capacitor can be removed again after the dynamic route arrives at point c. At the same time, the change in the inductor current is no longer negligible, and hence we must put the inductor back. In other words, after the dynamic route arrives at point c, we must go back to the original circuit in Fig. 20-3*a*.

The waveforms of $v_C(t)$ and $i_N(t) = -i_C(t)$ corresponding to the dynamic route in Fig. 20-3*e* are easily sketched as shown in Fig. 20-3*f* and *g*. The transition time T_r is easily seen to be given by $T_r = T_1 + T_2$, where

$$T_1 = -r_2 C \ln \frac{E_d - E_2}{E_b - E_2} \tag{20-10}$$

$$T_2 = r_3 C \ln \frac{E_d - E_3}{E_c - E_3} \tag{20-11}$$

The typical value of T_r is in the order of a few nanoseconds.

From Eqs. (20-10) and (20-11) we can draw the conclusion that the larger the value of the parasitic capacitance C, the larger the value of the transition time. Hence, it is impossible to accomplish an instantaneous jump ($T_r = 0$) unless $C = 0$.

The above example illustrates how a higher-order network may be reduced to a lower-order network during appropriate time intervals. In general, this technique can be applied to circuits having two or more time constants which differ greatly in their orders of magnitude. For example, the "coupling capacitor" used to couple the output signal from one black box into the input terminals of another black box can usually be replaced by a short circuit for purposes of analysis. Even if we are using a completely automatic computer program to analyze this class of networks, it is still important that we try to eliminate as many insignificant elements as possible. This will usually result in a great saving in computer time needed to analyze the network. In other words, it is important that we do not use the computer blindly; in certain problems, such as the example above, a heuristic analysis may be more satisfactory. For one thing, unless we are extremely careful, most computer programs would not anticipate the fast-changing waveform during the transition time T_r and would, therefore, give erroneous answers, unless the step size is programmed to vary in accordance with the rate of change of the waveforms.

---

[1] Recall the redundant series element property in Sec. 7.3.

20-5 THE QUALITATIVE APPROACH

We have already demonstrated in the preceding chapters the power of the qualitative approach. In general, as the exact solution becomes more difficult to obtain, the qualitative approach becomes more valuable. If the network is autonomous, the most common qualitative information that we seek is the location and stability of the equilibrium states associated with the network. The location of the equilibrium states can be easily found with the help of the explicit equilibrium state theorem discussed in Sec. 13-6. The stability of the equilibrium states can be determined by techniques beyond the scope of this book.

If the network is nonautonomous, then the qualitative information that we seek is usually the frequency-converting properties of the network. We have shown in Chap. 1 that one characteristic feature common to all nonlinear networks is the generation of new beat frequencies

$$\omega_{(m,n)} = m\omega_1 \pm n\omega_2 \tag{20-12}$$

by mixing two sinusoidal waveforms of frequencies ω_1 and ω_2.† In general, the higher the degree of nonlinearity, the more beat-frequency components are generated. Many applications in signal processing are based on this property alone. In all cases, we would *select* a number of frequency components from the many present by means of *filters.*[1] The important qualitative information that we seek in this case is the amount of power at a particular frequency that can be extracted.

To be more specific, consider the network shown in Fig. 20-4*a*. This is an *n*th-order nonautonomous network being driven by two sinusoidal sources of frequencies ω_1 and ω_2. Let us assume that

---

† For purposes of analysis, it is convenient to assume that the two frequencies ω_1 and ω_2 are *incommensurable,* i.e., their ratio is an *irrational number.*

[1] For our purpose here, a filter can be defined as a signal-processing black box whereby only a certain prescribed range of frequency components is allowed to reach the output terminals.

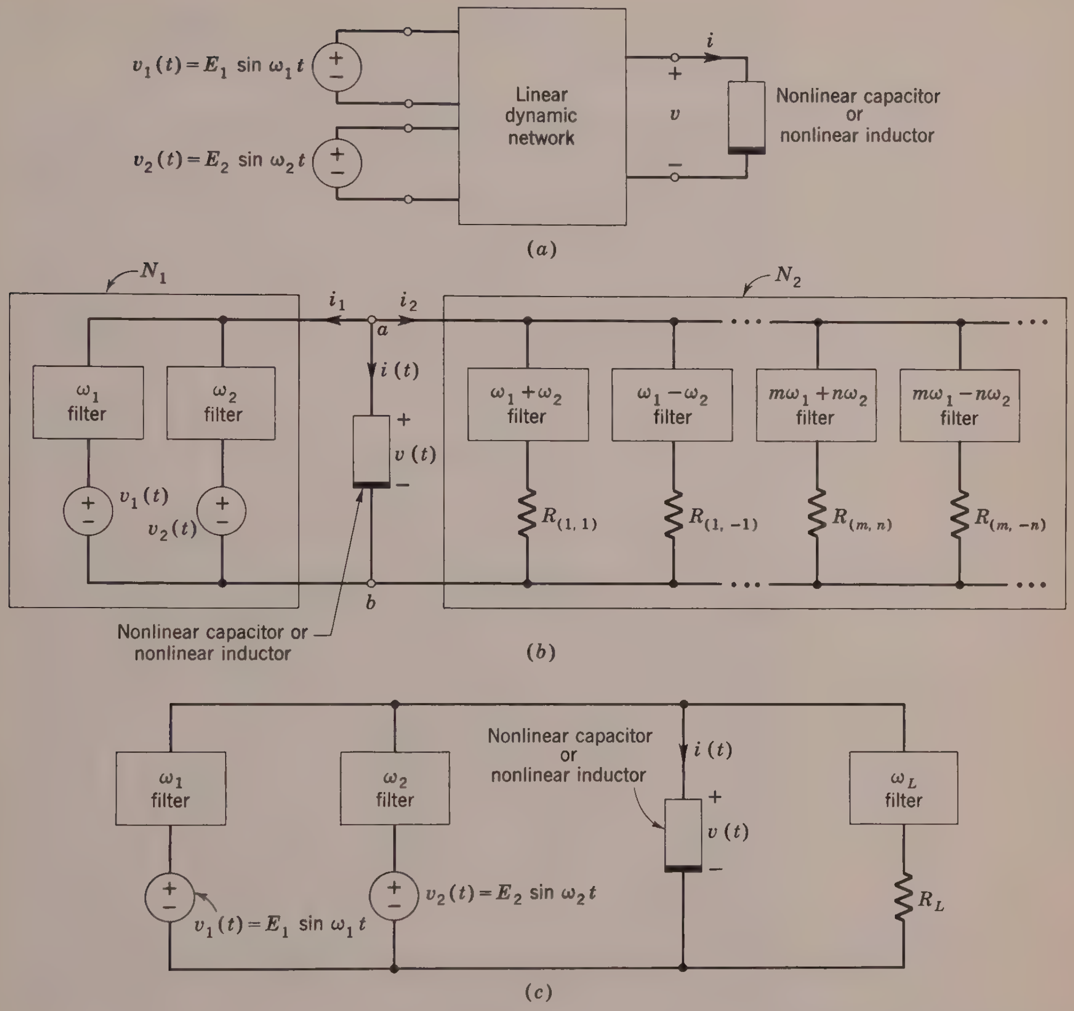

$v_1(t)=E_1 \sin \omega_1 t$
$v_2(t)=E_2 \sin \omega_2 t$
Linear dynamic network
i
$+$
v
$-$
Nonlinear capacitor or nonlinear inductor
(a)
N_1
N_2
i_1
i_2
a
$i(t)$
ω_1 filter
ω_2 filter
$v_1(t)$
$v_2(t)$
$v(t)$
$\omega_1+\omega_2$ filter
$\omega_1-\omega_2$ filter
$m\omega_1+n\omega_2$ filter
$m\omega_1-n\omega_2$ filter
$R_{(1,1)}$
$R_{(1,-1)}$
$R_{(m,n)}$
$R_{(m,-n)}$
b
Nonlinear capacitor or nonlinear inductor
(b)
ω_1 filter
ω_2 filter
Nonlinear capacitor or nonlinear inductor
$i(t)$
$v(t)$
ω_L filter
$v_1(t)=E_1 \sin \omega_1 t$
$v_2(t)=E_2 \sin \omega_2 t$
R_L
(c)

Fig. 20-4. An *n*th-order nonautonomous nonlinear network for illustrating the application of the Manley-Rowe equations.

the only nonlinear element in this network is either an inductor or a capacitor. The average power entering this nonlinear element is given by Eq. (1-45); namely,

$$P_{\text{av}} = \lim_{t\to\infty} \frac{1}{t_1} \int_0^t v(t)i(t)\,dt \tag{20-13}$$

For the sake of generality, let us suppose that both $v(t)$ and $i(t)$ contain all frequency components including a dc term corresponding to zero frequency. The instantaneous power $p(t) = v(t)i(t)$ will, upon expansion with the help of trigonometric identities, contain also all frequency components; namely,[1]

$$p(t) = p_{0,0} + p_{1,0} \cos \omega_1 t + p_{0,1} \cos \omega_2 t + p_{1,1} \cos (\omega_1 + \omega_2)t \\ + p_{1,2} \cos (\omega_1 + 2\omega_2)t + p_{2,1} \cos (2\omega_1 + \omega_2)t + \cdots \\ + p_{m,n} \cos (m\omega_1 + n\omega_2)t + \cdots \tag{20-14}$$

Substituting Eq. (20-14) into Eq. (20-13), we obtain

$$P_{\text{av}} = \sum_{m=0}^{\infty} \sum_{n=-\infty}^{\infty} P_{m,n} \tag{20-15}$$

where

$$P_{m,n} = \lim_{t\to\infty} \frac{1}{t_1} \int_0^t p_{m,n} \cos (m\omega_1 + n\omega_2)t \, dt \tag{20-16}$$

Hence, the total average power entering the nonlinear inductor or the nonlinear capacitor is the sum of the component average powers $P_{m,n}$ corresponding to the frequency $(m\omega_1 + n\omega_2)$. We have proved earlier that the average power entering any charge-controlled capacitor or any flux-controlled inductor is zero [see Eqs. (1-57) and (1-64)]. Hence, Eq. (20-15) becomes

$$\sum_{m=0}^{\infty} \sum_{n=-\infty}^{\infty} P_{m,n} = 0 \tag{20-17}$$

This relationship tells us only that the *total average power* is zero. The component powers $P_{m,n}$ are generally not zero. In order to investigate how the magnitudes of $P_{m,n}$ are distributed so that they cancel out each other in the end, let us *multiply* and *divide* each term $P_{m,n}$ by $(m\omega_1 + n\omega_2)$ and rewrite the equation as follows:

$$\omega_1\left[\sum_{m=0}^{\infty} \sum_{n=-\infty}^{\infty} \frac{mP_{m,n}}{(m\omega_1 + n\omega_2)}\right] + \omega_2\left[\sum_{m=0}^{\infty} \sum_{n=-\infty}^{\infty} \frac{nP_{m,n}}{(m\omega_1 + n\omega_2)}\right] = 0 \tag{20-18}$$

This equation must be true for any value of ω_1 and ω_2 and is independent of the shape of the nonlinearity of the energy-storage element. Since ω_1 and ω_2 are *incommensurable*, the terms inside the two brackets must independently be equal to zero; namely,[2]

$$\sum_{m=0}^{\infty} \sum_{n=-\infty}^{\infty} \frac{mP_{m,n}}{(m\omega_1 + n\omega_2)} = 0 \tag{20-19}$$

---

[1] For simplicity we have chosen to represent each component by a cosine term $p_{m,n} \cos (m\omega_1 + n\omega_2)t$.

[2] A rigorous proof of these classic relationships is given in J. M. Manley and H. E. Rowe, Some General Properties of Nonlinear Elements–Part I, General Energy Relations, *Proc. IRE,* vol. 44, pp. 904–913, July, 1956.

$$\sum_{m=0}^{\infty} \sum_{n=-\infty}^{\infty} \frac{nP_{m,n}}{(m\omega_1 + n\omega_2)} = 0 \tag{20-20}$$

These two relationships are known as the *Manley-Rowe equations.* They are very useful because they give valuable qualitative information concerning the network's capability. The usefulness of the Manley-Rowe equations is best exploited by considering the network shown in Fig. 20-4*b*, where a nonlinear energy-storage element is shown connected in parallel with the two black boxes labeled N_1 and N_2. Each rectangle inside N_1 and N_2 represents a filter, which is usually an *n*th-order dynamic linear network. Each filter is assumed to allow only currents with the designated frequency to flow through the corresponding resistor. All other frequency component currents are rejected.[1] If we apply KCL at node *a*, we find that as a consequence of the string of filters, the current $i(t)$, which would normally contain all frequency components, will now contain only those frequency components for which there is a filter.

As an application of the above observations, consider the special case shown in Fig. 20-4*c* where only three frequencies are allowed, namely, ω_1, ω_2, and $\omega_L \equiv M\omega_1 + N\omega_2$. The frequencies ω_1 and ω_2 correspond to the frequency of the two excitation sources, whereas the frequency ω_L corresponds to the frequency at the load resistor R_L. Hence, among the infinitely many possible frequency components, only the three components corresponding to ω_1, ω_2, and ω_L are not zero for this network; namely, $P_{1,0}(m = 1, n = 0)$, $P_{0,1}(m = 0, n = 1)$, and $P_{M,N}(m = M, n = N)$. Hence, the Manley-Rowe equations reduce to the following:

$$\frac{(1)P_{1,0}}{(1)(\omega_1) + (0)(\omega_2)} + \frac{(0)P_{0,1}}{(0)(\omega_1) + (1)(\omega_2)} + \frac{(M)P_{M,N}}{(M)(\omega_1) + (N)(\omega_2)} = 0 \tag{20-21}$$

$$\frac{(0)P_{1,0}}{(1)(\omega_1) + (0)(\omega_2)} + \frac{(1)P_{0,1}}{(0)(\omega_1) + (1)(\omega_2)} + \frac{(N)P_{M,N}}{(M)(\omega_1) + N(\omega_2)} = 0 \tag{20-22}$$

Solving for $P_{M,N}$ from Eqs. (20-21) and (20-22), we obtain

$$P_{M,N} = -\left(\frac{M\omega_1 + N\omega_2}{M\omega_1}\right) P_{1,0} \tag{20-23}$$

$$P_{M,N} = -\left(\frac{M\omega_1 + N\omega_2}{N\omega_2}\right) P_{0,1} \tag{20-24}$$

---

[1] This requirement defines an *ideal* filter. A practical filter can only attenuate but not completely eliminate the unwanted frequencies.

Bearing in mind that $P_{m,n} > 0$ implies power *entering* the energy-storage element, whereas $P_{m,n} < 0$ implies power *leaving* the energy-storage element,[1] we can now obtain the following interpretations.

Case 1 $M = 1$, $N = 1$. In this case the filter ω_L is tuned to a frequency $\omega_L = \omega_1 + \omega_2$, and Eqs. (20-23) and (20-24) become

$$P_{1,1} = -\left(\frac{\omega_1 + \omega_2}{\omega_1}\right) P_{1,0} \tag{20-25}$$

$$P_{1,1} = -\left(\frac{\omega_1 + \omega_2}{\omega_2}\right) P_{0,1} \tag{20-26}$$

Observe that these equations imply that $P_{1,1} < 0$ whenever $P_{1,0} > 0$ and $P_{0,1} > 0$. This means that while the energy-storage element is accepting power from the two sources at frequencies ω_1 and ω_2, it is delivering power simultaneously at frequency $(\omega_1 + \omega_2)$ to the load resistor. In practice, the energy-storage element is usually a varactor diode, and the excitation at frequency ω_1 is usually a very weak signal, whereas the excitation at frequency ω_2 is usually chosen so that $\omega_2 \gg \omega_1$. In this case, Eq. (20-25) can be written into the more convenient form

$$G \equiv \left|\frac{P_{1,1}}{P_{1,0}}\right| = \frac{\omega_1 + \omega_2}{\omega_1} = 1 + \frac{\omega_2}{\omega_1} \tag{20-27}$$

where G is called the *power gain* of the circuit. Observe that since $\omega_2 \gg \omega_1$, $G \gg 1$, and hence the output load power at frequency $\omega_L = (\omega_1 + \omega_2)$ is much higher than the excitation power at frequency ω_1. We have shown, therefore, that it is possible to design a power amplifier using a nonlinear energy-storage element such as a varactor diode. These networks are usually called *parametric amplifiers.*[2]

The above observation seems to suggest that we have violated the principle of conservation of energy, for how could we amplify power? The answer is that the source of the power comes from the second excitation source at frequency ω_2. Hence, the effect of the nonlinear energy-storage element is to convert the energy at frequency ω_2 into energy at frequency $(\omega_1 + \omega_2)$. Because of this interpretation, the excitation at frequency ω_2 is usually called the *pump*, whereas the excitation at frequency ω_1 is called the *signal.*

---

[1] Alternately, $P_{m,n} > 0$ implies that power is being *supplied* by the sources to the energy-storage element, whereas $P_{m,n} < 0$ implies that the energy-storage element is actually supplying the power to its external circuit.

[2] Since the load frequency $\omega_L = \omega_1 + \omega_2$ is greater than both ω_1 and ω_2, this type of network is also called an *up-converter.*

Case 2 $M = 1, N = -1$. In this case, the filter ω_L is tuned to a frequency $\omega_L = \omega_1 - \omega_2$, and Eqs. (20-23) and (20-24) become

$$P_{1,-1} = -\left(\frac{\omega_1 - \omega_2}{\omega_1}\right) P_{1,0} \tag{20-28}$$

$$P_{1,-1} = \left(\frac{\omega_1 - \omega_2}{\omega_2}\right) P_{0,1} \tag{20-29}$$

If we assume that $\omega_1 > \omega_2$, then Eq. (20-28) predicts that since $P_{1,0} > 0$ implies that $P_{1,-1} < 0$, power at the load frequency $\omega_L = \omega_1 - \omega_2$ is being supplied by the energy-storage element. However, since $(\omega_1 - \omega_2)/\omega_1 < 1$, the frequency ω_1 is being converted at a *loss,* and no power amplification is possible.

There are many other interesting observations that can be drawn from the Manley-Rowe equations. Observe that the main value of these equations lies in their complete independence of the network topology as well as of the shape of the nonlinear characteristic curve of the energy-storage element. In short, this qualitative result is truly remarkable for it tells so much for so little.

20-6 THE APPROXIMATE ANALYTICAL APPROACH

Beyond the methods and techniques described in the preceding sections, there exist a number of more mathematical and esoteric methods which we shall refer to as the *approximate analytical approach.*[1] Here it is imperative that the equations be obtained first in analytic form. Once this preliminary procedure is undertaken, a formidable task in itself, we may apply a number of nonlinear mathematical techniques which are beyond the scope of this book. The advanced readers who desire more information can find a list of references in Appendix E.

20-7 SUMMARY

There are generally five distinct approaches for analyzing higher-order nonlinear networks, which are the following:

1. Numerical-integration approach. Most amenable to the development of completely automatic user-oriented computer programs. The most serious problems at present are the numerical inaccuracies that tend to build up and the excessive amount of computer time needed to analyze relatively simple networks.

---

[1] Although these methods are analytical in nature, the solutions are invariably in the form of a series containing an infinite number of terms. In practice, only a few terms are kept, and even these require considerable effort; hence, the methods are really *approximate* rather than exact.

These difficulties are not fundamental, however, and hence this approach is certain to gain wide usage as more accurate and efficient computer programs become available.

2. Piecewise-linear approach. This method is based on the concept of the switching space and has certain advantages over the numerical-integration approach with respect to numerical accuracy and efficiency.
3. Heuristic approach. This approach is usually tailored for specific networks. It is based on intuition, experience, and a number of simplifying assumptions. Whenever applicable, the heuristic approach is generally superior to the other approaches.
4. Qualitative approach. Instead of seeking specific solutions corresponding to each initial state, this approach is concerned only with finding the general properties of the network. Some common properties of interest are the location and stability of equilibrium states in the case of autonomous networks. For nonautonomous networks the frequency-conversion properties of nonlinear elements are of much practical significance. The Manley-Rowe equations are invaluable for obtaining these properties.
5. Approximate analytical approach. This approach requires the normal-form equations to be derived in analytic form. Advanced mathematical backgrounds are generally required.

PROBLEMS

20-1 Describe a computer numerical-analysis algorithm for obtaining the solution of the network shown in Fig. 20-1*a* in terms of the following choice of state variables:
  (a) Capacitor charges and inductor flux linkages
  (b) Capacitor charges and inductor currents
  (c) Capacitor voltages and inductor flux linkages

20-2 Describe a computer piecewise-linear algorithm for obtaining the solution of the iterative piecewise-linear network shown in Fig. 20-2*a* for the following cases:
  (a) All energy-storage elements are linear.
  (b) All energy-storage elements are nonlinear but monotonic.
  (c) All capacitors are voltage-controlled, and all inductors are current-controlled.

20-3 Assuming that the parasitic capacitor in the tunnel-diode multivibrator circuit shown in Fig. 20-3*b* has a capacitance of $C = 10$ pF, find the exact waveforms of $v_L(t)$ and $i_L(t)$ during the transition time interval and calculate the transition time T_r.

20-4 Specialize the Manley-Rowe equations to the case where only the following frequency components are allowed:

(a) $\omega_1, \omega_2, \omega_1 + \omega_2, \omega_1 - \omega_2$

(b) $\omega_1, \omega_2, M\omega_1 + N\omega_2, M\omega_1 - N\omega_2$

(c) $\omega_1, \omega_2, \omega_1 + \omega_2, 2\omega_1 + \omega_2, \omega_1 + 2\omega_2, 2\omega_1 + 2\omega_2$

20-5 From the simplified Manley-Rowe equations obtained in Prob. 20-4, derive as many properties about the network as possible. Can you offer some practical applications using some of these properties?

20-6 Epic (*equilibrium point identification and classification*) is a recently developed computer program for finding the location and stability of the equilibrium states of an *autonomous* nth-order network. The algorithm is based on a theorem which states that a singular point $P(\bar{x}_1,\bar{x}_2, \ldots ,\bar{x}_n)$ of the nth-order system

$$\frac{dx_j}{dt} = f_j(x_1,x_2, \ldots ,x_n) \qquad j = 1,2, \ldots ,n$$

is *asymptotically stable* if the real parts of all eigenvalues of the *Jacobian matrix*

$$\mathbf{J}(x_1,x_2, \ldots ,x_n) \equiv \begin{bmatrix} \partial f_1/\partial x_1 & \partial f_1/\partial x_2 & \cdots & \partial f_1/\partial x_n \\ \partial f_2/\partial x_1 & \partial f_2/\partial x_2 & \cdots & \partial f_2/\partial x_n \\ \cdots & \cdots & \cdots & \cdots \\ \partial f_n/\partial x_1 & \partial f_n/\partial x_2 & \cdots & \partial f_n/\partial x_n \end{bmatrix}$$

evaluated at $x_1 = \bar{x}_1$, $x_2 = \bar{x}_2, \ldots , x_n = \bar{x}_n$ are negative and is unstable if the real part of at least one eigenvalue is positive.

(a) Show that each element of the Jacobian matrix corresponding to the network shown in Fig. 20-1*a* can be found by analyzing the resistive n-port subnetwork N under appropriate terminations.

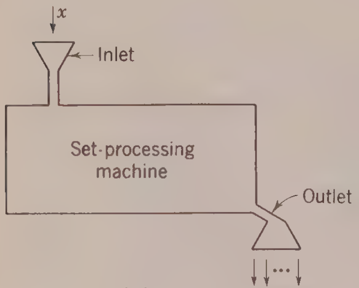
(b) Using Meca as a subprogram, construct a detailed flow chart for implementing Epic.

APPENDIX A MATHEMATICAL REPRESENTATION OF NONLINEAR NETWORK ELEMENTS

A-1 RELATIONS AND THEIR INVERSES

We shall be mainly concerned with definitions in this section. Just as in any dictionary, where a certain number of undefined words must be used in order to define other words, in mathematics we also have to use a number of undefined terms to define other terms. We shall only need one undefined term for our purpose here; namely, the word *set*. Intuitively, any collection of objects is called a set. Hence, we may talk about a set of letters (a, b, c, d, e), a set of animals (dogs, cats, goats), a set of v-i curves (Γ_1, Γ_2, Γ_3, Γ_4), a set of voltage waveforms [$v_1(t)$, $v_2(t)$, $v_3(t)$], a set consisting of all integers ($1, 2, \ldots, n, \ldots$), etc. Let us next introduce the notion of a *set-processing machine* having an inlet and an outlet as shown in Fig. A-1*a*. This machine has the property that whenever we dump an object x into its inlet, out comes one or more objects $y, z, \ldots, w$ through its outlet. For example, a slot machine might deliver one or more coins at its outlet corresponding to each coin (of a certain denomination) that enters its inlet. Depending on its internal structure, a set-processing machine can be identified by listing in a

Fig. A-1. Symbolic representation of a set-processing machine and a typical table describing it.



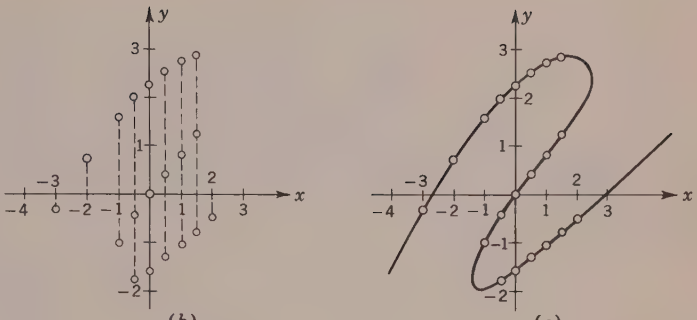
(*a*)

Input	Output			
x_1	y_1	z_1	$\cdots$	w_1
x_2	y_2	z_2	$\cdots$	w_2
$\vdots$	$\vdots$	$\vdots$	$\vdots\ \vdots\ \vdots$	$\vdots$

(*b*)

Domain x	Range y		
−3.0	−0.3		
−2.0	0.75		
−1.0	−1.0	1.6	
−0.5	−1.8	−0.4	2.0
0	−1.6	0	2.3
0.5	−1.3	0.4	2.5
1.0	−1.05	0.8	2.75
1.5	−0.8	1.25	2.9
2.0	−0.5		

(*a*)



(*b*) (*c*)

Fig. A-2. A relation between two sets of real numbers can be represented by a graph.

table the outputs corresponding to *each* input as shown in Fig. A-1*b*. It is important to observe that the machine may produce more than one output object for each input object that enters it. If we denote the set of all inputs by $D \equiv \{x_1, x_2, \ldots\}$ and the set of all outputs by $R \equiv \{y_1, y_2, \ldots, z_1, z_2, \ldots, w_1, w_2, \ldots\}$, then the machine is said to operate on a member of set D to produce one or more members of set R. In mathematics a set-processing machine is called a *relation,* and the sets D and R are called the *domain* and the *range* of the relation, respectively. A relation is, therefore, no more than a table that lists, corresponding to each member of its domain D, one or more members of its range R.

The simplest type of relation of interest to us corresponds to the case where both the domain and the range consist of real numbers. For example, the numbers listed in the table shown in Fig. A-2*a* are a relation. Observe that for certain values of x, the relation may have one, two, or three possible values for y. In the case of real numbers, it is more convenient to represent a relation in the form of a graph as shown in Fig. A-2*b*. Observe that the points in the graph represent exactly the same information as the numbers in the table. Aside from providing a simple geometrical representation, this graph has the advantage that it can list infinitely many points in the form of a curve, such as the relation shown in Fig. A-2*c*. Although equivalent conceptually, a table representing this relation is clearly impractical because it would require an infinite number of entries. In view of this graphical representation, any curve, or combination of curves or points, in the x-y plane is a relation. In fact, we may consider the terms "curve" and "relation" as being synonymous.

Let us return to our set-processing machine M once again and define an associated machine M^{-1} called the *inverse machine.* The purpose of this machine is to take each output member y of

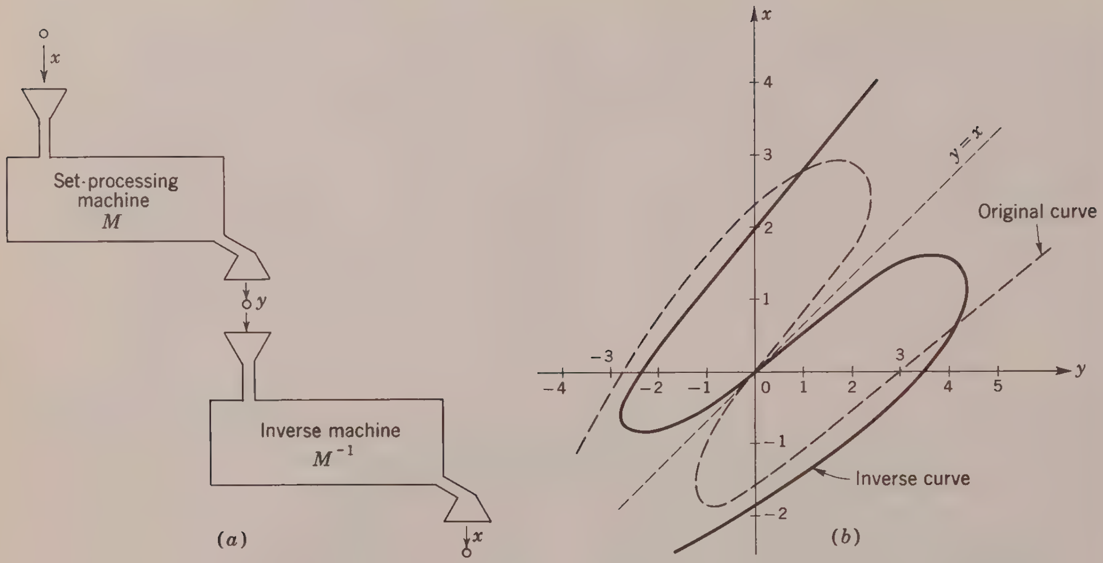


Fig. A-3. A symbolic representation of the relationship between a relation and its inverse.

M as its input, and to produce one or more outputs representing the set of inputs $x_1, x_2, \ldots$ of M which produced the original output y. The relationship between these two machines is best illustrated symbolically as shown in Fig. A-3*a*. Here the machine M takes an input object x from the domain of M and produces an output object y belonging to the range of M. We have assumed for simplicity that y is the only output in this case. If we dump y into the inverse machine M^{-1}, we shall get back the output x. In general, there may be other outputs in addition to x. However, for illustrative purposes, only the output x is shown. For obvious reasons, the relationship defining an inverse machine is called an *inverse relation.* Observe that an inverse relation M^{-1} is always defined with respect to an original relation M. The domain of M^{-1} is identical with the range of M. Conversely, the range of M^{-1} is the domain of M.

In the special case where both the domain and the range of M consist of the set of real numbers, the domain and range of M^{-1} will also be real numbers. Since the domain and range of M^{-1} correspond, respectively, to the range and domain of M, the graph representing M^{-1} can be simply obtained by interchanging the coordinates of each point on the original graph representing M. This is accomplished by *reflecting* the original graph with respect to the line $y = x$. For example, corresponding to the relation shown in Fig. A-2*c*, we obtain the inverse relation M^{-1} by reflecting the graph above the dotted line shown in Fig. A-3*b*. Observe that as mentioned previously, certain values of the input variable y may correspond to more than one value of the output variable x. A moment's inspection of Fig. A-3*b* reveals that if we make two reflections with respect to the same line, we shall obtain the original unreflected graph. This shows that the inverse of an inverse relation is the original relation.

A-2 FUNCTIONS AND THEIR INVERSES

In the definition of a set-processing machine and its inverse machine, we have seen that, corresponding to each input, one may get more than one output. This accounts for the fact that there is generally more than one column under the heading "output" (Figs. A-1*b* and A-2*a*). There is a special but extremely important class of relations wherein there is only one output corresponding to each input. In terms of the table defining the relation, this means that only one column appears under the output heading. A relation which satisfies this property is called a *function.* We shall denote a function by the letter f. Since for each member x belonging to the domain D of f there corresponds only one member y belonging to the range R of f, it is standard practice to represent this correspondence by the notation

$$y = f(x) \tag{A-1}$$

A simple example of a function might be that of a cigarette vending machine. In this case, the domain of the function f corresponds to the set of buttons for selecting the desired cigarettes, and the range of the function f corresponds to the set of different brands of cigarettes inside the machine. For each button x that we press, the function releases a particular brand of cigarettes $y = f(x)$. It is important to distinguish the difference between a function f and the output of the function $f(x)$ corresponding to an input x. The symbol f denotes the machine as a whole, but the symbol $f(x)$ denotes only one particular output of the machine.

If a relation M qualifies as a function, its inverse relation M^{-1} need not be a function because it is possible to have more than one input giving the same output. However, if the inverse relation M^{-1} is also a function, then M^{-1} is called an *inverse function.* If f is a function, we shall denote its inverse function, if it exists, by

$$x = f^{-1}(y) \tag{A-2}$$

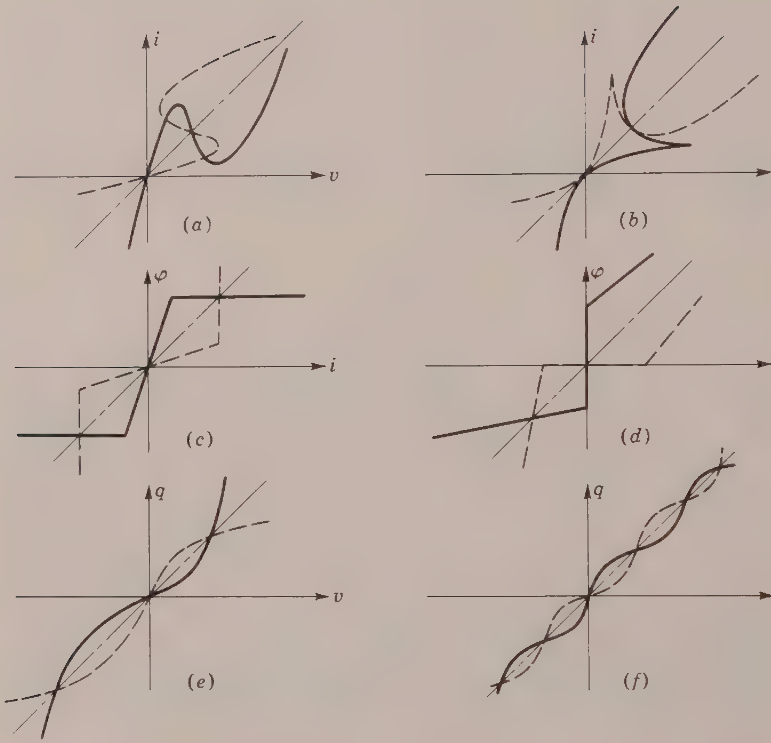


Fig. A-4. Examples illustrating the concept of a function and its inverse function.

A function f which has an inverse function f^{-1} is said to be a *one-to-one function.*

Specializing now to the x-y plane, it is clear that a graph representing a relation is a function if, and only if, every vertical line (representing a value of the domain) intersects the graph at one, and only one, point. The graph shown in Fig. A-2*c* does not satisfy this property and is, therefore, not a function. Instead of calling this a relation, we sometimes prefer to call it a *multivalued function.*

As an illustration of graphs which are functions, consider the curves shown in Fig. A-4, where the solid curves correspond to the given relations and the dotted curves correspond to the inverse relations. Observe that all solid curves except those shown in Fig. A-4*b* and *d* are functions. The v-i curve of Fig. A-4*b* is not a function because there are vertical lines which intersect this curve at more than one point. The i-φ curve of Fig. A-4*d* is also not a function because there are infinitely many values of φ corresponding to $i = 0$. However, the inverse relations of these two graphs are functions. On the other hand, the inverse relations of the two functions of Fig. A-4*a* and *c* are not functions. Finally, observe that only the last two graphs can be considered as one-to-one functions.

A common property is shared by both of these graphs; namely, $f(b) > f(a)$ whenever $b > a$. Such a curve is said to be *strictly monotonically increasing.* The solid curve shown in Fig. A-4c is monotonically increasing but *not strictly* because $f(b) = f(a)$ for some $b > a$. It is easy to see that a function is one-to-one if, and only if, it is characterized either by a strictly monotonically increasing curve, or a strictly monotonically decreasing curve. One-to-one functions are significant because they allow one to represent a curve mathematically as a function of either variable. This property is crucial in solving many nonlinear equations.

A function need not be restricted to the one-dimensional case. For example, we can consider the set of all m-tuples $(x_1,x_2, \ldots ,x_m)$ as the domain of the function and the set of all n-tuples $(y_1,y_2, \ldots ,y_n)$ as the range of the function. In this case, it is not possible to represent the function geometrically. However, it will still be convenient to represent this function symbolically as shown in Fig. A-5. Each point p in the domain D is assumed to have m coordinates $(x_1,x_2, \ldots ,x_m)$ and each point q in the range R is assumed to have n coordinates $(y_1,y_2, \ldots ,y_n)$. It is sometimes convenient to denote a function f having a domain D and a range R by the symbol

$$f\colon \quad D \rightarrow R \tag{A-3}$$

Usually, the domain consists of the entire m-dimensional space of real numbers R_m, and the range consists of the entire n-dimensional space of real numbers R_n. In this case, it is customary to write Eq. (A-3) in the form

$$f\colon \quad R_m \rightarrow R_n \tag{A-4}$$

Observe that the function f *maps* each point p in D into one, and only one, point q in R. Hence, q is usually referred to as the *image* of p. It is customary to denote this symbolically by

$$q = f(p) \tag{A-5}$$

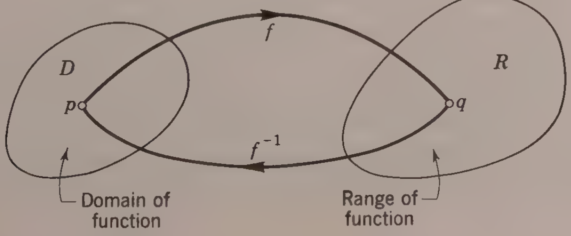


Fig. A-5. Symbolic representation of functions not describable by a plane curve.

If the function f satisfies the additional property that no two points in D map onto the same point in R, then a one-to-one correspondence exists between each point in D and each point in R. Such a function is, therefore, a one-to-one function and has an inverse. If we denote the inverse function by f^{-1} as shown in Fig. A-5, then each point q in R maps into one point p in D; namely,

$$p = f^{-1}(q) \tag{A-6}$$

The most common example of functions in higher-dimensional spaces is of the form

$$\begin{aligned} y_1 &= f_1(x_1,x_2, \ldots ,x_n) \\ y_2 &= f_2(x_1,x_2, \ldots ,x_n) \\ &\cdots\cdots\cdots\cdots \\ y_n &= f_n(x_1,x_2, \ldots ,x_n) \end{aligned} \tag{A-7}$$

If this function is also one-to-one, then each point $p(x_1,x_2, \ldots ,x_n)$ must correspond to only one point $q(y_1,y_2, \ldots ,y_n)$, and vice versa. Hence, Eq. (A-7) can also be written in its inverse form.

$$\begin{aligned} x_1 &= f_1^{-1}(y_1,y_2, \ldots ,y_n) \\ x_2 &= f_2^{-1}(y_1,y_2, \ldots ,y_n) \\ &\cdots\cdots\cdots\cdots \\ x_n &= f_n^{-1}(y_1,y_2, \ldots ,y_n) \end{aligned} \tag{A-8}$$

If the variables $x_1, x_2, \ldots, x_n$ in Eq. (A-7) are unknowns, then the one-to-one property implies that Eq. (A-6) will have a unique solution regardless of the value of $y_1, y_2, \ldots, y_n$.

A-3 MATHEMATICAL REPRESENTATION OF FUNCTIONS

There are many nonlinear network problems that cannot be solved by a purely graphical approach. In such cases it is sometimes desirable to obtain a mathematical representation of all characteristic curves which are given in graphical form. We shall consider in this section only the simpler case where the curves are functions and leave the multivalued case for the next section.

Although the theory of representation of functions is too advanced for our purpose here, there are a number of fundamental ideas and results that are simple enough to be presented. Two basic concepts are of particular interest to us, namely, those of *interpolation* and *approximation.* The problem of interpolation is concerned with finding a function

$$y = f(x) \tag{A-9}$$

which *passes through* a set of n prescribed points $P_1, P_2, \ldots, P_n$ in the x-y plane. Where the function goes in between these points is immaterial. Clearly, if $n = 2$, the function need only be a straight line. If $n = 3$, a parabola can always be found; if $n = 4$, a third-degree polynomial can always be found, etc. In fact, using the theory of determinants, it is easy to prove the following result.

FUNDAMENTAL INTERPOLATION THEOREM

Given a set of n points having distinct abscissas in the x-y plane, it is always possible to find one, and only one, $(n - 1)$-degree polynomial passing through these points.

The polynomial with the above property is an example of an *interpolation function.* A function made up of a sequence of linear interpolation functions is called a *piecewise-linear function.* Such a function can be represented mathematically by

$$y = m_j x + k_j \qquad \text{whenever } a_j \le x \le b_j \qquad j = 1, 2, \ldots \tag{A-10}$$

For convenience the linear function defined in each interval of Eq. (A-10) will be called a *segment,* and the end points of each segment will be called *breakpoints.* Clearly, a piecewise-linear function is the simplest representation of a function. It is easy to see that by increasing the number of segments, a piecewise-linear function can be used to represent a given function to within any degree of accuracy. A serious drawback of this representation is the fact that each breakpoint has two distinct slopes. Hence, a piecewise-linear function is not differentiable at a breakpoint. Moreover, since four numbers (m_j, k_j, a_j, and b_j) are required to describe each segment, it is clear that if the given function is highly nonlinear, other forms of representation may be simpler. The above considerations lead to the problem of approximation by differentiable functions.

Depending on the shape of the curve to be represented, a number of elementary functions may be chosen. For example, the *v-i* curve of a junction diode can be approximated fairly well by the following exponential function:

$$i = I_s[e^{(q/kT)v} - 1] \tag{A-11}$$

where the parameters I_s, q, K, and T are constants associated with a specific diode. Unfortunately, mother nature does not often give us devices characterized by curves that can be represented by elementary functions. To be sure, it is indeed rare to find that Eq. (A-11) does represent a junction-diode curve. Knowing now where we stand, it is natural to ask whether it is possible to find a set of "basic" functions and use them as building blocks to represent a given function. It would be too ambitious in any case to find a function that would pass through all points of a given curve. However, it is reasonable to hope that a function $g(x)$ might be found that will approximately represent a given curve $g(x)$ such that the error can be kept small. A classic theorem is available which says that this is indeed possible. Since the proof requires some background in mathematical analysis, we shall state only the result here.

WEIERSTRASS APPROXIMATION THEOREM

Let $g(x)$ be a continuous curve representing a function over some interval $a \leq x \leq b$. It is always possible to find a polynomial $p_n(x)$ of sufficiently high degree that the magnitude of the discrepancy between $p_n(x)$ and $g(x)$ is less than any prescribed positive number for all values of x within the interval (a,b).

Stated mathematically, this theorem guarantees that, provided we are willing to choose a polynomial of high enough degree, we can always find one such that

$$|p_n(x) - g(x)| < \epsilon \tag{A-12}$$

where ϵ is any arbitrarily small positive number. Observe that this theorem does not require the given curve $g(x)$ to be differentiable. The finite domain of $g(x)$ (namely, $a \leq x \leq b$) does not constitute a restriction in practice because the numbers a and b can be arbitrarily large, provided they are not infinite.

The Weierstrass approximation theorem guarantees only that the error ϵ can be made as small as we please (except zero).

It does not say whether the polynomial $p_n(x)$ will pass through a set of prescribed points as well. That this is also possible is given by the following theorem.

COMBINED INTERPOLATION AND APPROXIMATION THEOREM

A polynomial $P_n(x)$ of sufficiently high degree can always be found that satisfies not only the Weierstrass approximation theorem but also the additional requirement that it pass through an arbitrarily prescribed finite set of points having distinct abscissas.

To illustrate the above concepts, let us consider the v-i curve of a typical tunnel diode as shown in Fig. A-6*a*. First, let us find a polynomial interpolation function that will pass through the following four points on the curve: P_1 (0 volt, 0 ma), P_2 (0.15 volt, 1.0 ma), P_3 (0.7 volt, 0.06 ma), and P_4 (1.0 volt, 1.0 ma). Applying the fundamental interpolation theorem, we know there is a unique

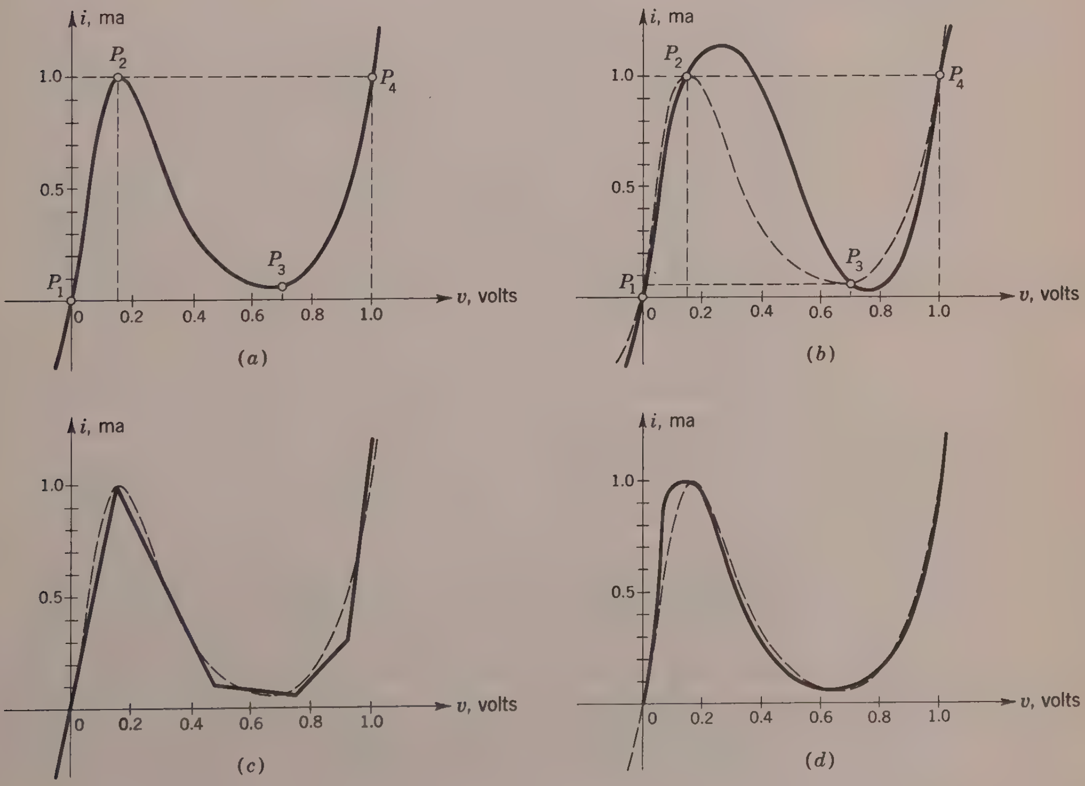


Fig. A-6. A typical tunnel-diode curve can be represented either by a polynomial or by a piecewise-linear function.

third-degree polynomial that will pass through these four points; namely,

$$p_3(x) = a_0 + a_1x + a_2x^2 + a_3x^3 \tag{A-13}$$

To solve for the coefficients a_0, a_1, a_2, and a_3, we substitute the coordinates (v,i) of each point P_j into Eq. (A-13), thus obtaining

$$\begin{aligned} 0 &= a_0 + a_1(0) + a_2(0)^2 + a_3(0)^3 \\ 1.0 &= a_0 + a_1(0.15) + a_2(0.15)^2 + a_3(0.15)^3 \\ 0.06 &= a_0 + a_1(0.7) + a_2(0.7)^2 + a_3(0.7)^3 \\ 1.0 &= a_0 + a_1(1) + a_2(1)^2 + a_3(1)^3 \end{aligned} \tag{A-14}$$

Equation (A-14) constitutes four linear equations in four unknowns. We can solve for these unknowns and obtain $a_0 = 0$, $a_1 = 10.10$, $a_2 = -26.50$, and $a_3 = 17.40$. The desired interpolation function is, therefore, given by

$$i = 10.10\,v - 26.50\,v^2 + 17.40\,v^3 \tag{A-15}$$

If we plot this function in the v-i plane as shown by the solid curve in Fig. A-6*b*, we find that it passes through the four prescribed points, as it should. Observe, however, that this function is a very poor representation of the given curve (shown dotted). We could have obtained a much better approximation by choosing five piecewise-linear segments as shown in Fig. A-6*c*. The resulting piecewise-linear function is easily seen to be represented by

$$i = \begin{cases} 6.67\,v & -\infty \le v \le 0.15 \\ -2.682\,v + 1.4073 & 0.15 \le v \le 0.48 \\ -0.2\,v + 0.21594 & 0.48 \le v \le 0.75 \\ 1.35\,v - 0.94656 & 0.75 \le v \le 0.92 \\ 10\,v - 8.90456 & 0.92 \le v \le \infty \end{cases} \tag{A-16}$$

Observe that Eq. (A-16) requires a total of 19 numbers to completely specify the function. Since the curve given in Fig. A-6*a* is not overly complicated, a higher-degree polynomial might fit this curve just as well, if not better. If we do not insist on the curve passing through a certain number of points, but only require that the error defined by Eq. (A-12) be small, we can find many polynomials that will have this property.

One such polynomial is given by

$$i = 17.76\,v - 103.79\,v^2 + 229.62\,v^3 - 226.31\,v^4 + 83.72\,v^5 \quad \text{(A-17)}$$

The curve represented by Eq. (A-17) is shown in Fig. A-6*d*. A comparison of this with the original curve (shown dotted) shows the approximation to be just as good as that shown in Fig. A-6*c*. However, this time only five parameters are needed. If a computer were being used, this function would take up a lot less memory. Another advantage of Eq. (A-17) is, of course, that it is differentiable everywhere. On the other hand, if the given curve is highly nonlinear, as in the case where there are many local maxima and minima, then a piecewise-linear representation may be simpler to specify. However, the real motivation for using piecewise-linear representations lies in the fact that it permits a highly nonlinear problem to be solved as several related linear problems.

The preceding discussions are concerned only with the function of one variable. There are many occasions where we have to represent a function of two or more variables, such as $z = f(x,y)$. Unfortunately, the interpolation theorem cannot be generalized to this case. It is possible, however, to extend the Weierstrass approximation theorem to functions of two or more variables.

A-4 MATHEMATICAL REPRESENTATION OF MULTIVALUED FUNCTIONS

The DP plot of a resistive black box containing two or more nonmonotonic resistors is usually a multivalued function. In order to study networks containing such multivalued elements analytically, it is necessary to obtain a mathematical representation of a multivalued curve. Clearly, it is not possible to represent such a curve by Eq. (A-9) since, by the very definition of a function, only one value of y is allowed for each value of x. We shall now introduce a new approach, the parametric approach, for representing any multivalued curve which satisfies the property of *unicursality* as defined below.

Any curve Γ in the x-y plane is said to be unicursal if, starting with one end of Γ (possibly at $-\infty$), it is possible to trace the entire curve in one continuous stroke without lifting the tip of the pencil from the paper and without retracing any portion of the curve.

Three examples of unicursal curves are shown in Fig. A-7*a* to *c*. The curves shown in Fig. A-7*d* to *f* are *not* unicursal because it is not possible to trace any one of these curves in a single con-

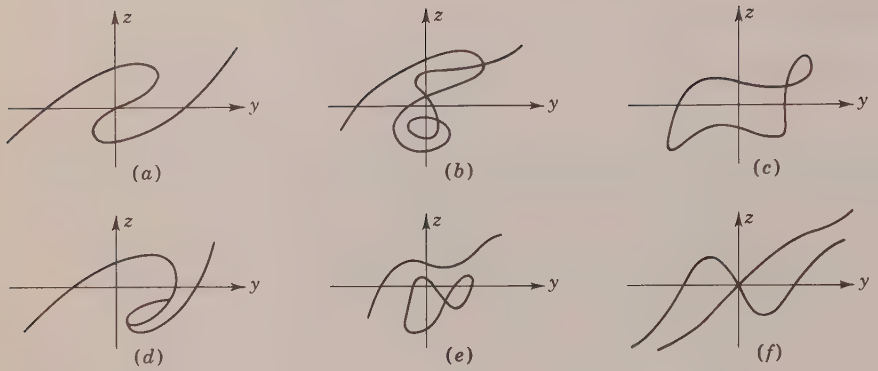


Fig. A-7. A unicursal curve may be (*a*) simple or (*b*) self-intersecting, or it may (*c*) form a closed loop. Not unicursal are curves (*d*) forming a self-loop, (*e*) having separate parts, or having more than two branches at infinity.

tinuous stroke. From these examples it should be clear that if one has an infinitely long piece of string, then a curve Γ is unicursal if, and only if, it is possible to lay down the string on a plane and form an exact facsimile of Γ. One property of an unstretchable string is that we can mark off any convenient unit of length from an arbitrary point on the string in the same way a tape measure for tailoring clothes is marked. If we do this and trace the curve Γ with the marked string once again, then this time each point on the curve is associated with a unique number, namely, its length from the origin.

In the case of the unicursal curve, we can arbitrarily choose an origin on the curve and mark off the *arc length* from this origin in both directions of the curve. To distinguish these two directions, we may arbitrarily mark the length along one of the two directions with a negative sign. If we denote the arc length measured from the origin along the curve by the parameter x, then it is clear that for each value of x there corresponds one, and only one, point $P(y,z)$ on the curve. Going back to our set-processing machine once again, we find that if we "dump" a value of x into the machine's inlet, we shall get one, and only one, point from the machine's outlet. In other words, the correspondence between the arc length and the points along a unicursal curve can be interpreted as a function. Since each point has two coordinates y and z, each coordinate may be represented as a function of the parameter x; namely

$$y = f(x) \tag{A-18}$$

and

$$z = g(x) \tag{A-19}$$

Equations (A-18) and (A-19) completely characterize the curve because upon performing a graphical elimination procedure, we can recover the original curve Γ. Let us demonstrate this with a few examples.

EXAMPLES

1. Consider the simple unicursal curve shown in Fig. A-8*a*. Choosing arbitrarily the point 0 as the origin and a positive direction as indicated by the arrow sign, we mark off the arc lengths in both directions, as shown in Fig. A-8*a*. (This can be done by straightening the curve and measuring its length.)

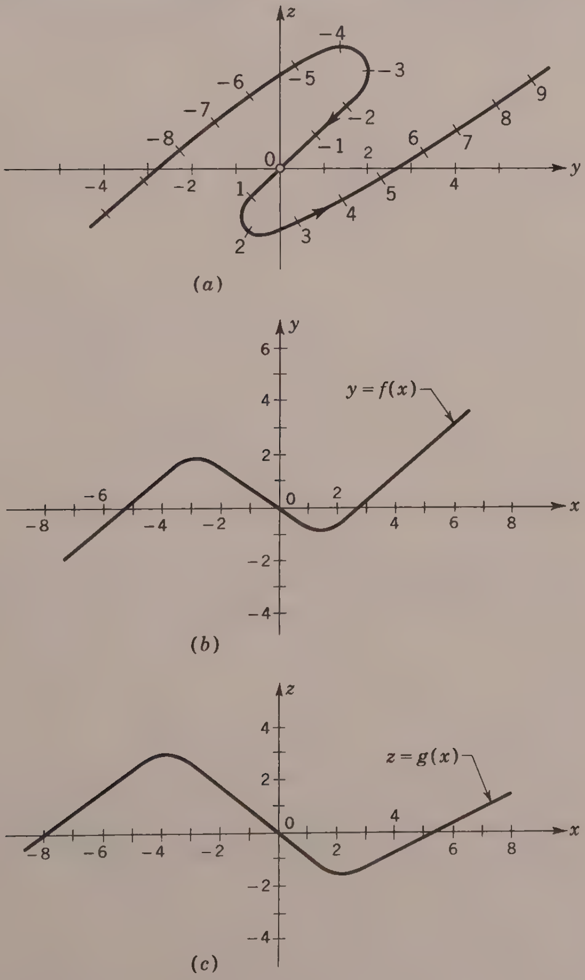


Fig. A-8. Parametric representation of a simple unicursal curve.

From the calibrated curve we obtain the two functions $y = f(x)$ and $z = g(x)$ as shown in Fig. A-8*b* and *c*, respectively. As expected, both curves come out to be functions. If necessary, we can now use the techniques of the preceding section to represent these curves by polynomials.

2. Consider next a more complicated unicursal v-i curve as shown in Fig. A-9*a*. For convenience we have represented this curve by piecewise-linear segments. With the origin 0 and the positive direction as indicated, the arc length for this case can be computed directly. The resulting parametric representations $v = v(x)$ and $i = i(x)$ are readily obtained as shown in Fig. A-9*b* and *c*.

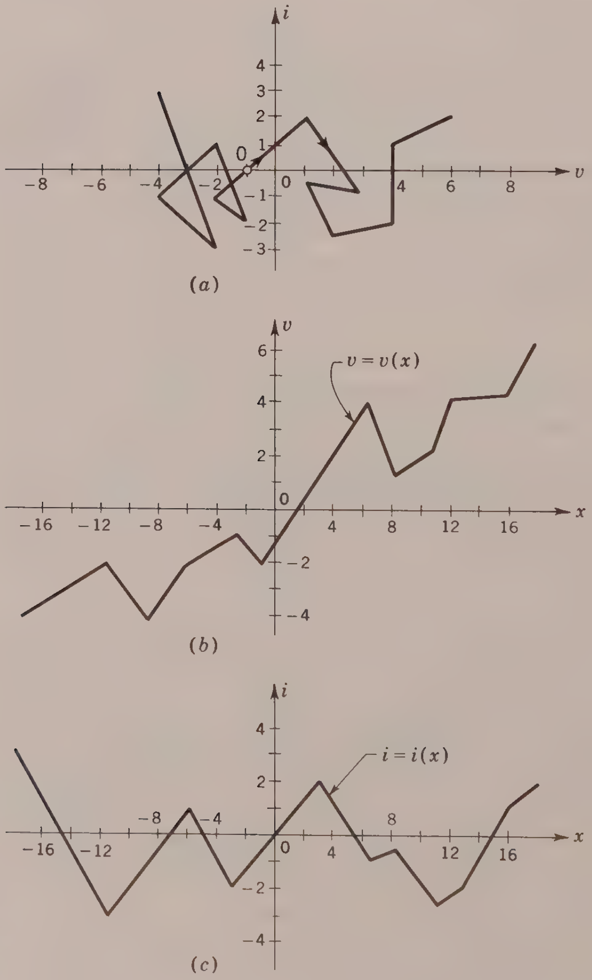


Fig. A-9. Parametric representation of a self-intersecting piecewise-linear unicursal curve.

A moment's inspection of Fig. A-9 shows that whenever the unicursal curve is piecewise-linear, the parametric curves will also be piecewise-linear. However, in any case observe that the choice of the origin is completely arbitrary. In fact, a simple exercise will show that the effect of changing the origin is merely to translate the parametric curves along the x axis.

A-5 PARAMETRIC REPRESENTATION OF NONLINEAR NETWORK ELEMENTS

We shall now apply the approach presented in the preceding section to represent nonlinear network elements characterized by unicursal curves. Such elements will henceforth be called *unicursal elements.*

A unicursal resistor is represented by two parametric functions,

$$v = v(x) \tag{A-20}$$

and

$$i = i(x) \tag{A-21}$$

A unicursal inductor is represented by two parametric functions,

$$i = i(x) \tag{A-22}$$

and

$$\varphi = \varphi(x) \tag{A-23}$$

Since $v = d\varphi/dt$, we can apply the chain rule to obtain the parametric expression for the inductor voltage

$$v = \frac{d\varphi(x)}{dx}\frac{dx}{dt} \tag{A-24}$$

A unicursal capacitor is represented by two parametric functions

$$v = v(x) \tag{A-25}$$

and

$$q = q(x) \tag{A-26}$$

Since $i = dq/dt$, we can apply the chain rule to obtain the parametric expression for the capacitor current

$$i = \frac{dq(x)}{dx}\frac{dx}{dt} \tag{A-27}$$

For convenience we have so far been referring to the parameter x as the arc length of the unicursal curves. However, this just happens to be a simple parameter to choose. There is no reason why other parameters cannot be chosen so long as the functions completed specify the unicursal curve. For example, suppose a unicursal resistor happens to be voltage-controlled. Then clearly it is simpler in this case to choose the voltage itself as a parameter, and Eqs. (A-20) and (A-21) reduce trivially to

$$v = x \tag{A-28}$$

$$i = i(x) = i(v) \tag{A-29}$$

Equation (A-28) is, of course, superfluous in this case. By similar reasoning, it is clear that with an appropriate choice of parameters, the parametric representation reduces to the conventional representation whenever the elements can be represented by functions. Therefore, the conventional approach is a special case of the parametric approach. This means that any theorem or procedure formulated in terms of the parametric approach is always applicable to conventional networks.

Consider next a network containing n unicursal elements. In this case, each element j is associated with a parameter x_j. Accordingly, the equations of motion obtained by applying KCL and KVL will all be expressed in terms of the parameters $x_1, x_2, \ldots, x_n$. We have, therefore, seen that using the parametric approach, a much larger class of nonlinear networks can be represented. Let us consider a specific example.

EXAMPLE

Consider the single-loop *RLC* network shown in Fig. A-10*a*. Let the nonlinear resistor R be characterized by the four-cusped hypocycloid shown in Fig. A-10*b*. This unicursal curve can be represented by

$$v_1 = \cos^3 x_1 \tag{A-30}$$

and

$$i_1 = \sin^3 x_1 \tag{A-31}$$

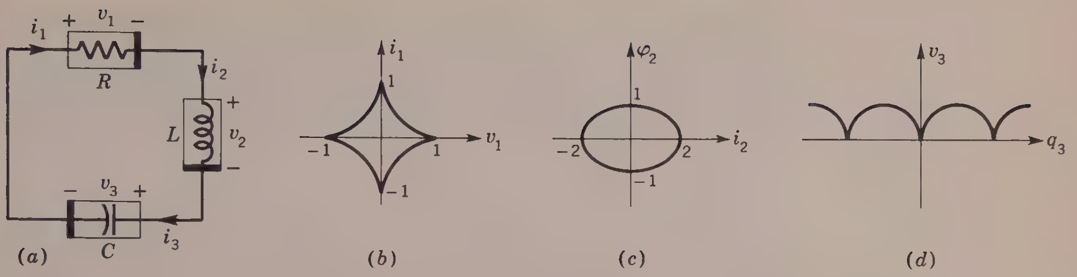


Fig. A-10. An illustrative unicursal nonlinear network.

Let the nonlinear inductor L be characterized by the ellipse shown in Fig. A-10*c*. This unicursal curve can be represented by

$$i_2 = 2 \sin x_2 \tag{A-32}$$

$$\varphi_2 = \cos x_2 \tag{A-33}$$

and

$$v_2 = \frac{d\varphi_2}{dx_2}\frac{dx_2}{dt} = -\sin x_2 \frac{dx_2}{dt} \tag{A-34}$$

Let the nonlinear capacitor C be characterized by the cycloid shown in Fig. A-10*d*. This unicursal curve can be represented by

$$v_3 = 1 - \cos x_3 \tag{A-35}$$

$$q_3 = x_3 - \sin x_3 \tag{A-36}$$

and

$$i_3 = \frac{dq_3}{dx_3}\frac{dx_3}{dt} = (1 - \cos x_3)\frac{dx_3}{dt} \tag{A-37}$$

To write the equations of motion, we obtain from KCL and KVL the following equations:

$$i_1 = i_2 \tag{A-38}$$

$$i_3 = i_2 \tag{A-39}$$

$$v_1 + v_2 + v_3 = 0 \tag{A-40}$$

Substituting into these equations the respective currents and voltages given by Eqs. (A-30) to (A-37), we obtain

$$\sin^3 x_1 = 2 \sin x_2 \tag{A-41}$$

$$(1 - \cos x_3)\frac{dx_3}{dt} = 2 \sin x_2 \tag{A-42}$$

$$\cos^3 x_1 - \sin x_2 \frac{dx_2}{dt} + (1 - \cos x_3) = 0 \tag{A-43}$$

Equations (A-41) to (A-43) constitute three nonlinear functional-differential equations in the three unknowns x_1, x_2, and x_3. The next step is obviously to solve these equations. Since our purpose here is merely to demonstrate how the equations of motion can be written in parametric form, we shall not proceed to find the solutions. It suffices, however, to say that once the solutions $x_1(t)$, $x_2(t)$, and $x_3(t)$ are found, the voltages and currents of each element can be easily determined from Eqs. (A-30) to (A-37).

In Chap. 1 we restricted our discussion of the energy flow into an energy-storage element to that of a flux-controlled inductor and charge-controlled capacitor. We shall now apply the parametric approach to derive similar expressions for the more general case of unicursal inductors and capacitors. Let us start with the unicursal relation for energy flow into a one-port black box; thus,

$$w(t_0,t_1) = \int_{t_0}^{t_1} i(t)v(t)\,dt \tag{A-44}$$

For a unicursal inductor we can substitute Eq. (A-22) and (A-23) in place of i and v in Eq. (A-44) and obtain

$$w_L(t_0,t_1) = \int_{t_0}^{t_1} i(x(t))\,\frac{d\varphi(x)}{dx}\,\frac{dx}{dt}\,dt \tag{A-45}$$

By a standard change of variable, Eq. (A-45) reduces to

$$w_L(t_0,t_1) = \int_{x(t_0)}^{x(t_1)} i(x)\,\frac{d\varphi(x)}{dx}\,dx \tag{A-46}$$

In the special case where the inductor is flux-controlled, we can simply choose the parameter $x = \varphi$, in which case Eq. (A-46) reduces Eq. (1-61) in Chap. 1.

By a dual procedure the energy flow entering a unicursal capacitor is given by

$$w_C(t_0,t_1) = \int_{x(t_0)}^{x(t_1)} v(x)\,\frac{dq(x)}{dx}\,dx \tag{A-47}$$

Again, observe that upon choosing $x = q$ for a charge-controlled capacitor, Eq. (A-47) reduces to Eq. (1-54) in Chap. 1.

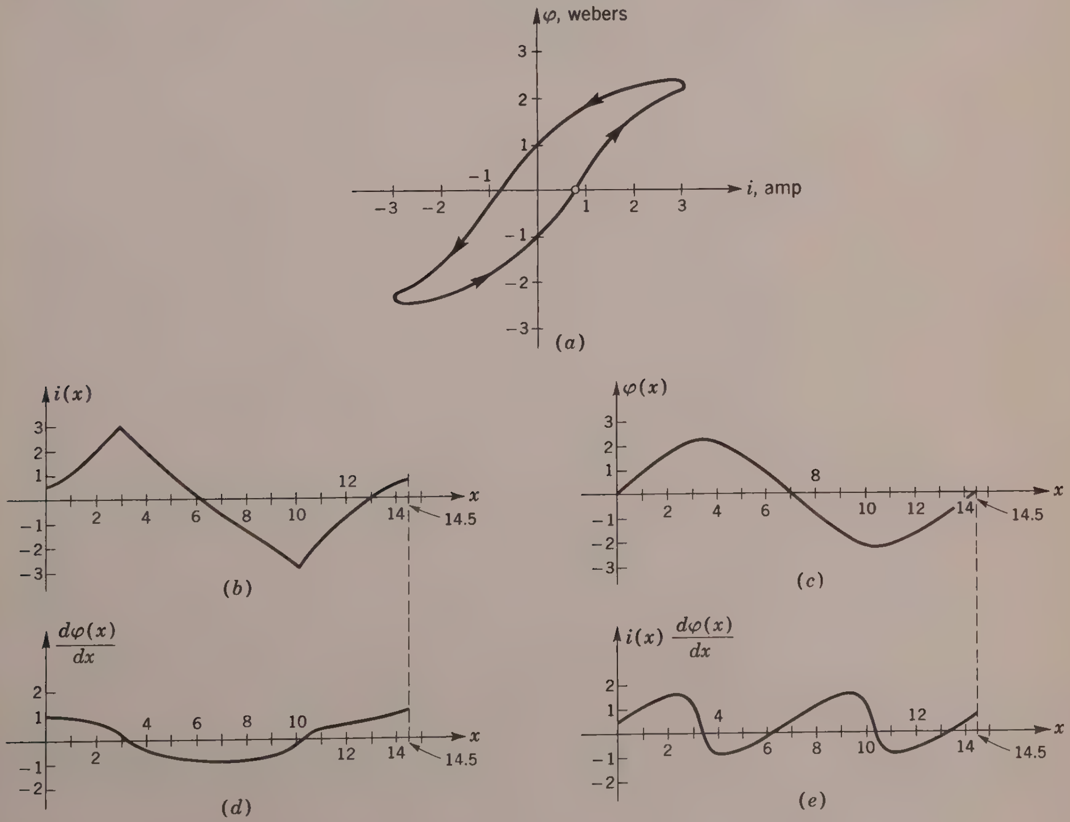
An examination of Eqs. (A-46) and (A-47) shows that both $w_L(t_0,t_1)$ and $w_C(t_0,t_1)$ depend only on the value of the parameter x corresponding to the initial point at $t = t_0$ and the value of the parameter x corresponding to the terminal point at $t = t_1$. The exact nature of the waveforms of $v(t)$ and $i(t)$ is immaterial. From this observation, we can conclude by the same reasoning as that used in Chap. 1 that, provided the unicursal curve does not intersect itself, the average power will be zero. Such inductors and capacitors can only store energy and are, therefore, lossless. This conclusion is *not* true if the unicursal curve intersects itself, thus forming a closed loop. This is because each time one goes around a closed loop and returns to the original point, the value of x increases by an amount equal to the perimeter of the loop. Therefore, even though we may be back at the same point at which we started,

$$x(t_1) \neq x(t_0) \tag{A-48}$$

and the energy entering the element from t_0 to t_1 is not returned to the external circuit. If we apply a periodic signal to such an element, a finite amount of energy is lost or dissipated per cycle, and the average energy cannot be zero. Such an inductor or capacitor is, therefore, not lossless.

Let us verify the preceding result by considering a physical example. Suppose we apply a periodic voltage across a nonlinear inductor with hysteresis. Although no valid technique is presently available for describing the hysteresis phenomenon, our assumption of a periodic signal guarantees that the inductor will be operating around a closed i-φ loop as shown in Fig. A-11a. Since

Fig. A-11. A demonstration that the energy lost per cycle is equal to the area enclosed by the hysteresis curve.



a closed loop is unicursal, we can obtain its parametric representations $i = i(x)$ and $\varphi = \varphi(x)$ as shown in Fig. A-11*b* and *c*, respectively. Observe that these parametric functions are periodic; hence, only one period is shown. To apply Eq. (A-46), we must obtain $d\varphi(x)/dx$ and multiply it by $i = i(x)$. This can be done by the graphical differentiation and multiplication techniques presented in Appendix B. The results are shown in Fig. A-11*d* and *e*. If we graphically integrate Eq. (A-46) over one period, then since $x(t_0) = 0$ and $x(t_1) = 14.5$, even though they seem to be the same point, we obtain

$$w_L(t_0,t_1) = \int_0^{14.5} i(x)\frac{d\varphi(x)}{dx}\,dx = 5.8 \text{ watts} \tag{A-49}$$

Equation (A-49) shows that 5.8 watts of energy is being dissipated per cycle in this case. From experience we know that this energy is lost in the form of heat generated in the iron core of the inductor. In fact, it has been verified *experimentally* that this energy loss per cycle is equal to the area enclosed by the hysteresis loop. If we can determine the area of the loop shown in Fig. A-11*a* by a planimeter or by any other means, we would find the area to be approximately 5.8 watts, as it should be.

APPENDIX B GRAPHICAL TECHNIQUES FOR BASIC MATHEMATICAL OPERATIONS

In the absence of a complete theory in nonlinear mathematics, it is usually impractical, if not impossible, to solve a nonlinear problem analytically. Because of this, two basic approaches are generally used for solving such problems. They are *numerical techniques* and *graphical techniques.* The first approach is extremely powerful and general when used with a computer. Since many books have been written on numerical techniques, we shall not consider them here. The second approach is less general, but it has at least four significant advantages over the numerical approach. The most important advantage is that it gives the user considerable insight into the overall nature of the problem. In fact, on many occasions the solution of one problem by graphical techniques automatically leads to the solution of several related problems with very little additional effort. The second advantage is that graphical techniques are generally valid even in the case of multivalued problems.[1] The third advantage is that all curves involved in the problem are specified in graphical form. This feature is very desirable in practice because the characteristics of most nonlinear systems are invariably obtained empirically and hence are not represented in analytic form. The fourth advantage of graphical techniques is that, whenever applicable, they are usually simple to understand because they seldom require any mathematical background beyond calculus. This feature is particularly ideal for our purpose, and we have taken full advantage of it in this book.

Graphical techniques usually involve the successive application of a sequence of six basic mathematical operations. They are (1) addition and subtraction, (2) multiplication and division, (3) composition, (4) elimination, (5) differentiation, and (6) integration. The first four involve operations between two or more functions, whereas the last two involve operations on one function only. Our purpose in this section is to present a simple procedure for implementing these operations graphically.

---

[1] This is not true in numerical techniques because the assumption of uniqueness is almost always required.

B-1 GRAPHICAL ADDITION AND SUBTRACTION

Consider any two functions defined by

$$y_1 = f_1(x_1) \tag{B-1}$$

and

$$y_2 = f_2(x_2) \tag{B-2}$$

Let y_3 be the sum of y_1 and y_2; that is,

$$y_3 \equiv f_3(x_1,x_2) = f_1(x_1) + f_2(x_2) \tag{B-3}$$

This sum can be obtained graphically by considering either x_1 or x_2 as a parameter, thereby reducing Eq. (B-3) to the sum of a function and a constant. Therefore, for each value of the assumed parameter, Eq. (B-3) represents a vertical translation of either $f_1(x_1)$ or $f_2(x_2)$. For example, if $f_1(x_1)$ and $f_2(x_2)$ are given, respectively, as shown in Fig. B-1*a* and *b*, then their sum is given by the system of curves shown in Fig. B-1*c* with x_2 chosen as the parameter. In the special case where the variables $x_1 = x_2 = x$, the curves in Fig. B-1*c* can be reduced to one curve by obtaining the value of y_3 at the intersection between the curve with parameter $x_2 = x$ and the vertical line $x_2 = x$, for each value of x. The result is shown in Fig. B-1*d*. Of course, in this case it is easier to obtain the sum directly by aligning the axis y_1 and y_2 and then adding the ordinates corresponding to each value of $x = x_1 = x_2$.

Exactly the same procedure applies in the case of subtraction. In fact, the two curves need not even be functions, as in the case of multivalued curves. In this case the procedure must be applied to each branch of the curves separately, and all possible combinations must be included.

B-2 GRAPHICAL MULTIPLICATION AND DIVISION

Consider Eqs. (B-1) and (B-2) again, but this time let

$$y_3 = f_3(x_1,x_2) = f_1(x_1)f_2(x_2) \tag{B-4}$$

This product can be obtained graphically by considering either x_1 or x_2 as a parameter, thereby reducing Eq. (B-4) into the product

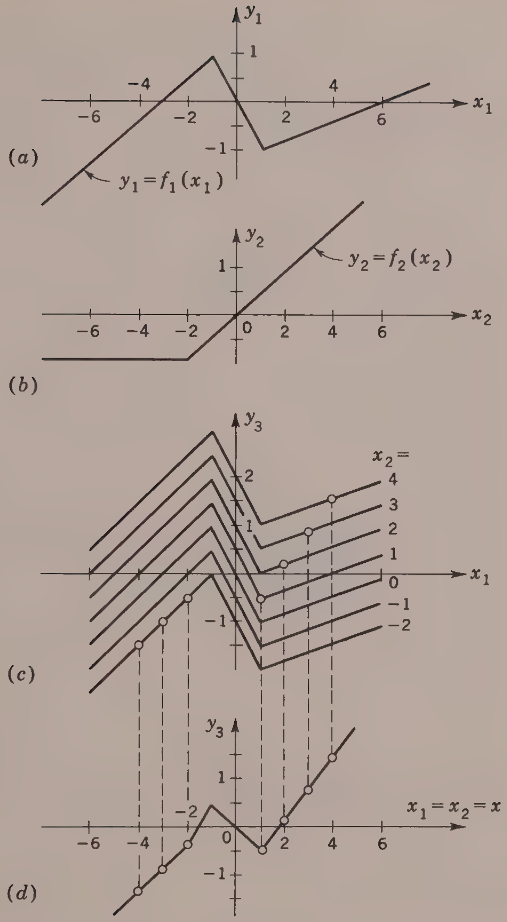


Fig. B-1. Graphical procedure for adding two functions.

of a function and a constant. For example, if $f_1(x_1)$ and $f_2(x_2)$ are given, respectively, as shown in Fig. B-2*a* and *b*, then the product is given by the system of curves shown in Fig. B-2*c* with x_1 chosen as the parameter. In the special case where $x_1 = x_2 = x$, the curves in Fig. B-2*c* can be reduced to one curve by the same procedure as before. The result is shown in Fig. B-2*d*. Of course, in this case, it is easier to align the two vertical axes and multiply the ordinates corresponding to the same value of x directly.

B-3 GRAPHICAL COMPOSITION

Suppose we are given two functions

$$z = f(y) \tag{B-5}$$

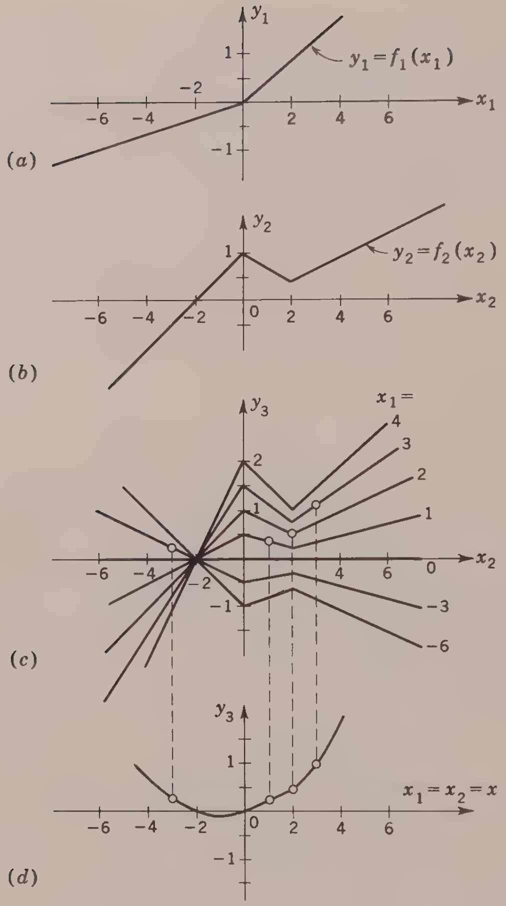


Fig. B-2. Graphical procedure for multiplying two functions.

and

$$y = g(x) \tag{B-6}$$

If we substitute Eq. (B-6) in place of y in Eq. (B-5), we obtain a new function

$$z = h(x) = f(g(x)) \tag{B-7}$$

This process of taking the function of another function is called the *composition* operation. It is clear from Eq. (B-7) that the composition operation always leads to a new function.

A simple graphical procedure can be used to obtain the composition between two functions. This consists of plotting on quadrille paper the two functions $z = f(y)$ and $y = g(x)$ in the relative positions shown in Fig. B-3. The new function $z = h(x)$ is located in the upper left-hand corner. Since the domain of the new function is x, the procedure consists of dropping vertical guidelines for each value of x until it intersects the curve $y = g(x)$, at which point it is then projected horizontally to the right. Upon hitting the unity slope line, the guideline continues upward until it intersects the curve $z = f(y)$, at which point it is projected horizontally to the left. The intersection of the two guidelines in the z-x plane gives one point of the new function $z = h(x)$. This procedure is then repeated for different values of x. Obviously, if drawn on quadrille paper, the guidelines need not be actually constructed. It is easy to see now that as long as $f(y)$ and $g(x)$ are functions, there will always be only one point in the x-z plane corresponding to each value of x. In other words, $z = h(x)$ is always a function. If either $f(y)$ or $g(x)$ or both are multivalued, the composition procedure is still valid. However, the resulting curve in the x-z plane will no longer be a function, and Eq. (B-7) must be replaced by a relation

$$h(x,z) = 0 \tag{B-8}$$

B-4 GRAPHICAL ELIMINATION

Suppose we are given two functions with the same independent variable y, namely,

$$z = f(y) \tag{B-9}$$

and

$$x = g(y) \tag{B-10}$$

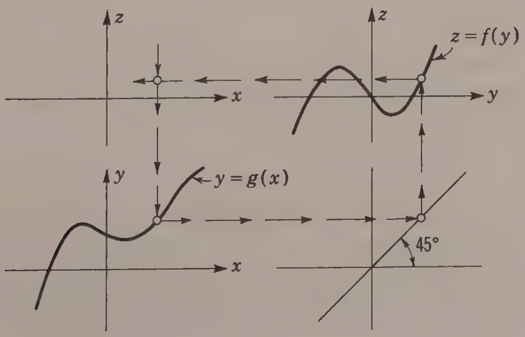


Fig. B-3. Graphical procedure for taking the composition of two functions.

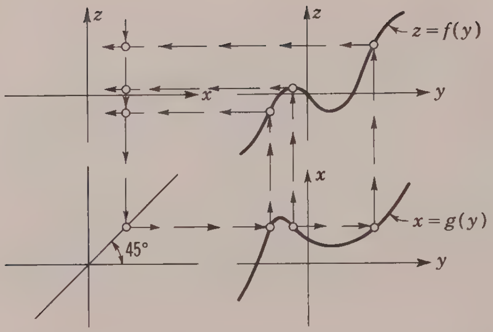


Fig. B-4. Graphical procedure for eliminating the common variable between two functions.

A common operation calls for eliminating the "common" variable y, thereby obtaining a new "relation"

$$h(x,z) = 0 \tag{B-11}$$

in the x-z plane. Observe that we have used the term "relation" instead of "function" because the resulting curve need not be a function. In fact, it is easy to see that in order for the result to be a function, we must take the inverse of Eq. (B-10), $y = g^{-1}(x)$, and substitute this in place of y in Eq. (B-9) to obtain the composition $z = h(x)$. But we already know that the inverse $y = g^{-1}(x)$ will be a function if, and only if, $x = g(y)$ is either a strictly monotonically increasing function, or a strictly monotonically decreasing function. Hence, the same condition is required for Eq. (B-11) to be a function. Observe that the elimination operation is equivalent to taking the composition between a function and the inverse of another function.

A simple graphical procedure can be used to eliminate the common variable between two functions. This is shown in Fig. B-4. Observe that the only difference between Figs. B-3 and B-4 is in the position of the unity slope line. Otherwise, the graphical constructions are exactly identical. A moment's inspection of Fig. B-4 reveals that if the curve $x = g(y)$ is not strictly monotonically increasing, or strictly monotonically decreasing, the lower horizontal projection line will intersect it at more than one point, thereby resulting in a multivalued function $z = h(x)$.

B-5 GRAPHICAL DIFFERENTIATION

Consider an arbitrary point Q on a curve $y = f(x)$ as shown in Fig. B-5a. Our purpose here is to find a simple graphical procedure for determining the slope at point Q. Suppose we draw a tangent line through Q until it intersects the x axis at point A. The slope at point Q is, therefore, given by

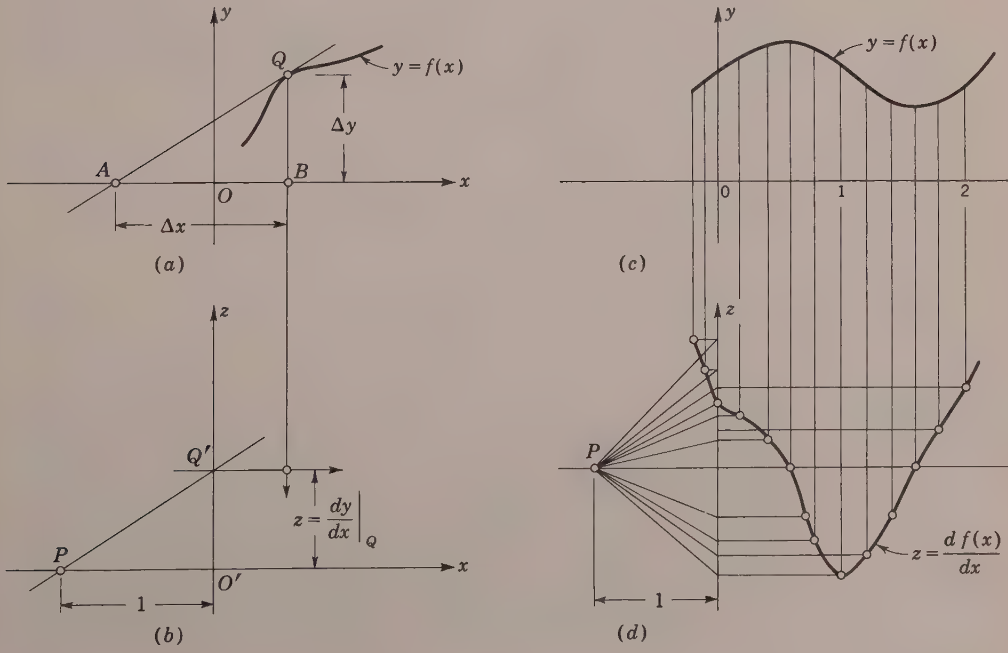
$$\left.\frac{dy}{dx}\right]_Q = \frac{\Delta y}{\Delta x} \tag{B-12}$$

Next, draw the x-z axes in alignment with the y axis as shown in Fig. B-5*b* and locate the point P, called the *pivot,* at a *unit* distance to the left of the origin. Through this point P draw a straight line parallel to the tangent line QA in Fig. B-5*a*. Let Q' be the intersection of this line with the z axis. A moment's inspection shows that the two triangles QAB and $Q'PO'$ are similar. Therefore,

$$\frac{Q'O'}{O'P} = \frac{Q'O'}{1} = \frac{dy}{dx} \tag{B-13}$$

Equation (B-13) shows that the z coordinate of point Q' in Fig. B-5*b* automatically gives the slope of point Q in Fig. B-5*a*. Therefore, the intersection between the horizontal line through Q' and the vertical line through Q establishes the derivative of $y = f(x)$ at Q. With a pair of triangles this procedure can be easily imple-

Fig. B-5. Graphical procedure for obtaining the derivative of a curve.



mented to obtain the derivative curve $z = dy/dx$. For example, applying this procedure to the curve shown in Fig. B-5*c*, we obtain the desired derivative curve shown in Fig. B-5*d*. Observe that the graphical construction could have been made directly on the *x*-*y* plane.

B-6 GRAPHICAL INTEGRATION

Consider first the simpler case of a staircase curve $y = f(x)$ as shown in Fig. B-6*a*. We would like to derive a simple graphical procedure for obtaining the integral

$$z(x) = \int_{x_0}^{x} f(x)\,dx \tag{B-14}$$

as a function of the upper limit of integration x. In view of the stepwise nature of $f(x)$, the integral of $f(x)$ will be piecewise-linear and will have a breakpoint Q_j at each value $x = x_j$ where $f(x)$ jumps abruptly. Hence, if the slope of each piecewise-linear segment can be determined graphically, then the integral of $f(x)$ can be readily obtained. To find this slope, it suffices to consider a pair of adjacent breakpoints Q_j and Q_{j+1} and the segment connecting them, as shown in Fig. B-6*a*. If we denote the value of the integral at Q_j and Q_{j+1} by $z(x_j)$ and $z(x_{j+1})$, respectively, then the increase in $z(x)$ from x_j to x_{j+1} must be equal to the area under the rectangle with a height $f(x_j)$; that is,

$$\Delta z \equiv z(x_{j+1}) - z(x_j) = (x_{j+1} - x_j)f(x_j) \tag{B-15}$$

From Eq. (B-15) we obtain the slope m_j of segment j.

$$m_j \equiv \frac{z(x_{j+1}) - z(x_j)}{x_{j+1} - x_j} = f(x_j) \tag{B-16}$$

Hence the slope of segment j is equal numerically to the height $f(x_j)$ of the stepwise curve at $x = x_j$. Now, suppose we locate a *pivoting point* P at a unit distance from the origin and draw a straight line connecting P and the point of intersection Q'_j between the y axis and the projection of the horizontal step $y = f(x_j)$. Observe that the slope of the line Q'_jP is equal to $f(x_j)$ and is, therefore, equal to the slope of segment j. This means that if we draw a line through Q_j parallel to Q'_jP until it cuts the vertical line $x = x_{j+1}$, the resulting segment is exactly the integral curve from

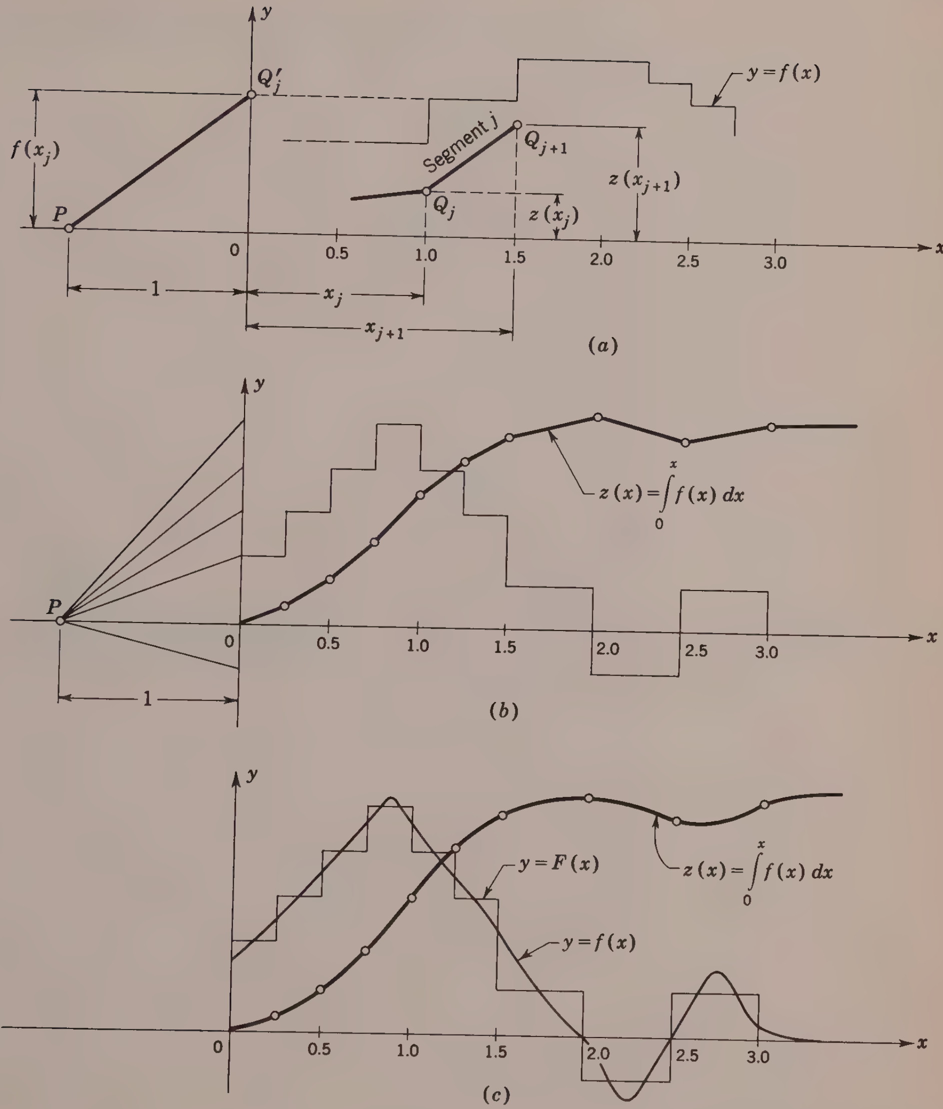


Fig. B-6. Graphical procedure for integrating a curve $y = f(x)$ from x_0 to x.

x_j to x_{j+1}. The terminal point at $x = x_{j+1}$ is the breakpoint Q_{j+1}. Starting from this point all over again, we can continue until the entire integral curve $z(x)$ is obtained. Figure B-6*b* shows an example of how this procedure is carried out. Observe that the solution obtained in this manner is exact. In practice, if the same vertical scale is chosen, the integral curve might become too large and thus leave the paper. In this case, we can easily change the scale of $z(x)$

by changing the distance of the pivoting point P. Clearly, if the distance PO is greater than unity, the ordinate of $z(x)$ will decrease, thereby effecting a change in scale.

When the curve $y = f(x)$ is not a stepwise function as is usually the case, we can devise a fairly accurate but approximate method, provided the curve $f(x)$ is sufficiently smooth. Under this assumption, the integral will also be smooth. This means that if we can accurately determine a string of points separated by small increments in x, then by drawing a smooth curve through these points, we should obtain a reasonably accurate solution curve. In order to obtain the string of points accurately, we can replace the curve $y = f(x)$ by a stepwise function $F(x)$ chosen such that the net area between the $F(x)$ and $f(x)$ over each interval approaches zero. For smoother curves this can be easily done visually. For example, the smooth curve $y = f(x)$ shown in Fig. B-6*c* might be approximated by the stepwise function $F(x)$. For convenience we have chosen $f(x)$ in Fig. B-6*c* so that the stepwise function $F(x)$ coincides with that shown in Fig. B-6*b*. Therefore, the string of points are identical as shown in Fig. B-6*c*. The final integral is obtained by passing a smooth curve through these points. Observe that this curve is a much better approximation of $z(x)$ than the piecewise-linear curve of Fig. B-6*b*.

APPENDIX C SCHEMATIC DIAGRAMS OF SOME BASIC CIRCUITS

C-1 CIRCUITS FOR TRACING v-i, v-q, AND i-φ CURVES

A simple circuit for tracing the v-i curve of a nonlinear *resistor* is shown in Fig. C-1. The Variac (variable autotransformer) is used to vary the amplitude of the sinusoidal input voltage, and the isolation transformer is used to isolate the common terminal of the tracer circuit from the chassis ground. The resistor R is used to monitor the current through the nonlinear resistor and may be assigned any convenient value if the v-i curve is monotonic. However, if the v-i curve is *voltage-controlled,* as in the case of a tunnel diode, the value of R must be chosen small enough so that at any time t the load line representing R will intersect the v-i curve at only one point. If the v-i curve is *current-controlled,* as in the case of a neon bulb, the value of R must be chosen large enough so that at any time t the load line representing R will intersect the v-i curve at only one point. It is important that the stray inductance and stray capacitance be kept as small as possible in the nonmonotonic case by using short leads. Otherwise, the circuit may become unstable and burst into oscillation.

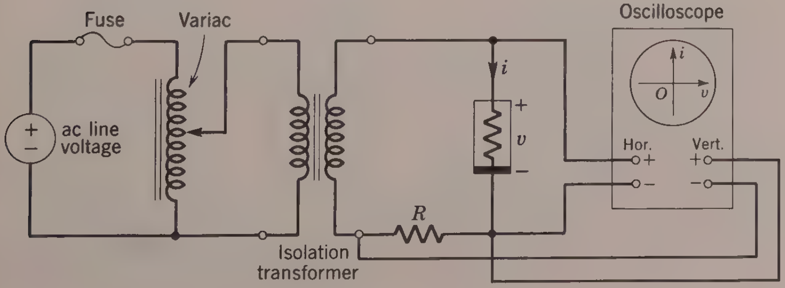


Fig. C-1. Schematic diagram of a v-i curve tracer.

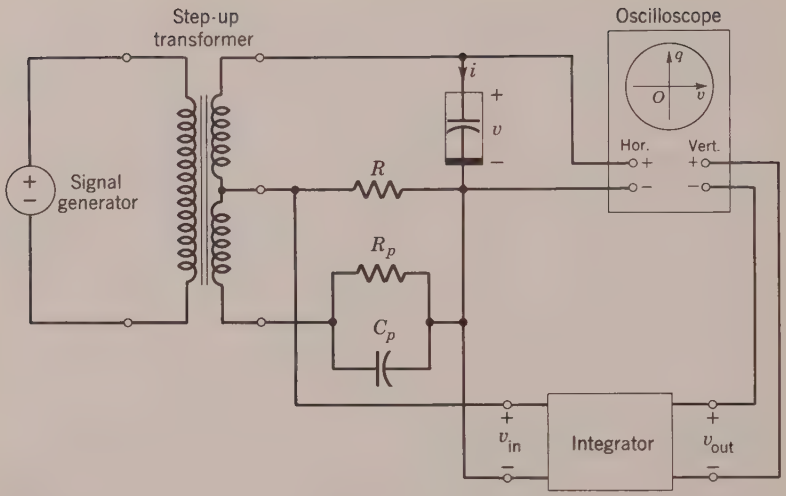


Fig. C-2. Schematic diagram of a v-q curve tracer.

To trace the v-q curve of a nonlinear *capacitor,* we must integrate the capacitor current in order to obtain the charge q. Similarly, to trace the i-φ curve of a nonlinear *inductor,* we must integrate the inductor voltage in order to obtain the flux linkage φ. A simple circuit for tracing the v-q curve of a nonlinear capacitor is shown in Fig. C-2. The dual circuit for tracing the inductor i-φ curve is shown in Fig. C-3. A signal generator is used in order to obtain a variable-frequency voltage source. The step-up transformer in either circuit is not absolutely necessary. However, there are two advantages in using the transformer. First, it is used to match the impedance of the tracer circuit to the generator (for this purpose the transformer may be replaced by a transistorized booster circuit such as an emitter-follower). A second advantage

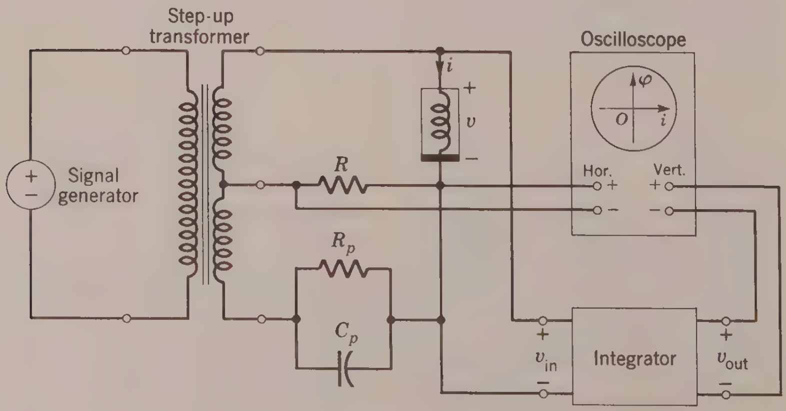


Fig. C-3. Schematic diagram of an i-φ curve tracer.

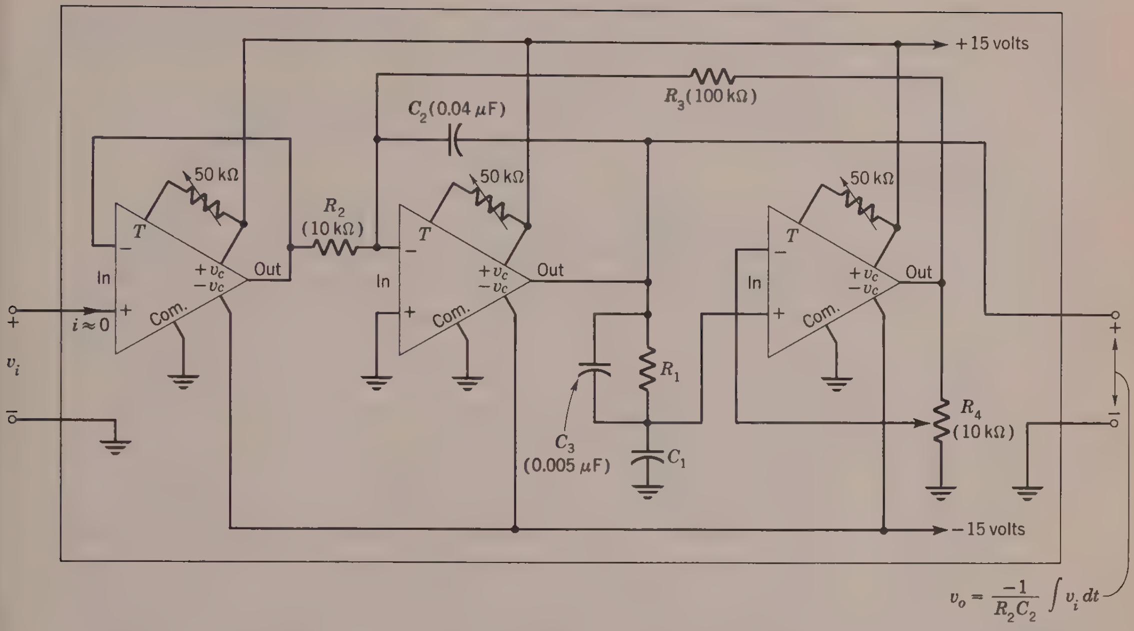


Fig. C-4. Schematic diagram of an integrator circuit for the curve tracers.

of using a transformer with a tapped secondary winding is that the input impedance and the stray capacitance of the oscilloscope can be neutralized by adjusting R_p and C_p. If, for example, the transformer is center-tapped, C_p should be set equal to the stray capacitance, and R_p should be set equal to the oscilloscope impedance. If the turns ratio of the bottom-half secondary winding to the upper-half secondary winding is k, then C_p should be set equal to $1/k$ times the stray capacitance, and R_p should be set equal to k times the oscilloscope input impedance. The current-monitoring resistor R should be on the order of 100 Ω for tracing v-q curves with incremental capacitance under 1 μF, and about 10 Ω if the incremental capacitance is above 1 μF. For the inductor curve tracer, the value of R should be about 100 Ω.

The integrator circuit designed for the above curve tracers is shown in Fig. C-4.[1] In order to minimize loading effects, an *isolation box* (see Fig. 3-20*b* of Chap. 3) is chosen to be the input stage. The second stage is the main integrating circuit (see Fig. 3-23*a*). In order to prevent the integrator from drifting into saturation, the third stage provides a negative feedback loop for eliminating the dc component of the integrator. The values for C_1 and R_1 should be chosen equal to 1 μF and 30 kΩ, respectively, for operating frequencies above 150 Hz. For lower frequencies, choose $C_1 = 10$ μF and $R_1 = 5$ MΩ.

---

[1] A Nexus SQ10A (or any equivalent) operational amplifier may be used for this circuit, as well as for the following circuits.

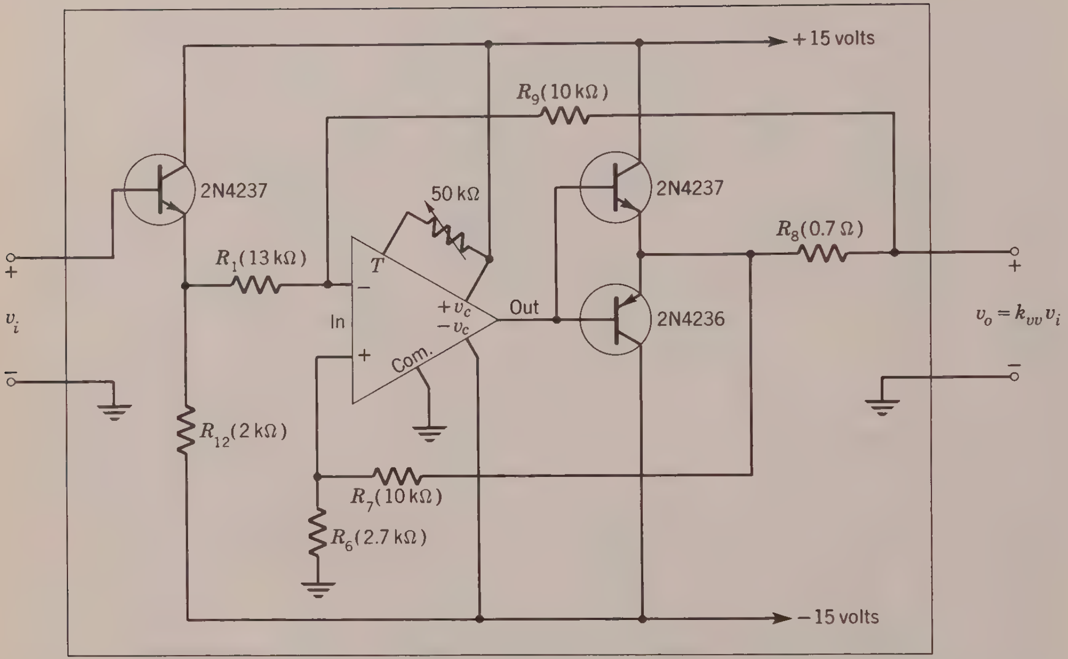


Fig. C-5. Schematic diagram of a voltage-controlled voltage source.

C-2 CIRCUITS FOR SIMULATING CONTROLLED SOURCES

The schematic diagrams for the four types of controlled sources are shown in Figs. C-5 to C-8. Observe that, except for the input stage, the circuit for the voltage-controlled voltage source in Fig. C-5 is identical with the right-half portion of the current-controlled voltage source in Fig. C-6. Observe also that except for the additional current-sensing resistor R_2, the circuit for the current-controlled current source in Fig. C-7 is identical with that of the voltage-controlled current source in Fig. C-8. Hence, with the help of switches, one needs only to build two basic circuits in order to simulate any one of the four types of controlled sources. The values of the resistors in these diagrams are given for a particular value of the controlling coefficients; namely, $k_{vv} \approx -1$ for the voltage-controlled voltage source, $k_{cv} \approx -1.3$ volts/ma for the current-controlled voltage source, $k_{cc} \approx 1$ for the current-controlled current source, and $k_{vc} \approx 10$ ma/volts for the voltage-controlled current source. A different controlling coefficient can be obtained by an appropriate choice of resistor values. A more detailed description of each circuit and the procedure for adjusting the controlling coefficients are given below.

Voltage-controlled voltage source To a first approximation, the controlling coefficient k_{vv} of the circuit shown in Fig. C-5 is given by

$$k_{vv} = \frac{v_o}{v_i} \approx \frac{-R_9}{R_1 - R_6[(R_1 + R_9)/(R_6 + R_7)]} \tag{C-1}$$

For satisfactory operation, R_8 should be less than an ohm to ensure high-current output, and R_{12} must be chosen small enough in order to maintain a proper bias on the emitter-follower transistor 2N4237. To ensure a stable operation, R_1 must be chosen greater than $R_6[(R_1 + R_9)/(R_6 + R_7)]$.

With the resistor values shown in Fig. C-5, the circuit can deliver a maximum current of ±0.1 amp. The output impedance is slightly negative ($-0.1\ \Omega < Z_0 < 0$) so that the circuit will cancel the lead resistance of the external circuit. The frequency response of this circuit is a function of the output voltage level. At 20 volts peak-to-peak (and within the allowable output current level) the response is flat to about 40 kHz. The phase shift between the input and output becomes significant above 20 kHz. At 2 volts peak-to-peak the response is flat to about 100 kHz. The phase

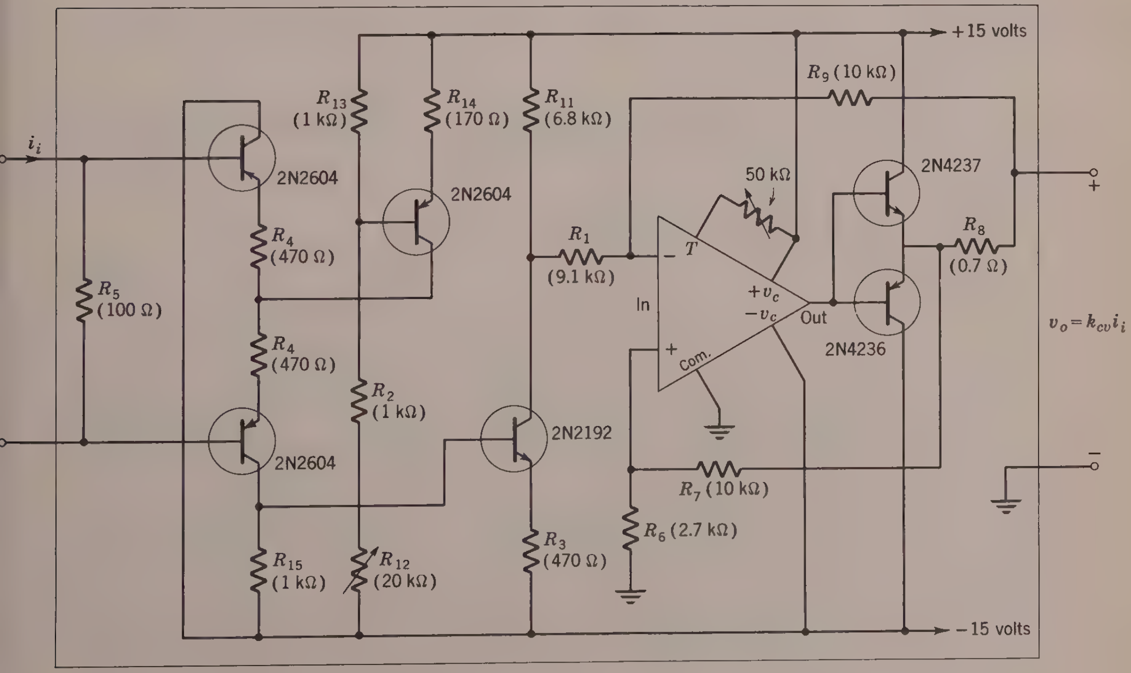


Fig. C-6. Schematic diagram of a current-controlled voltage source.

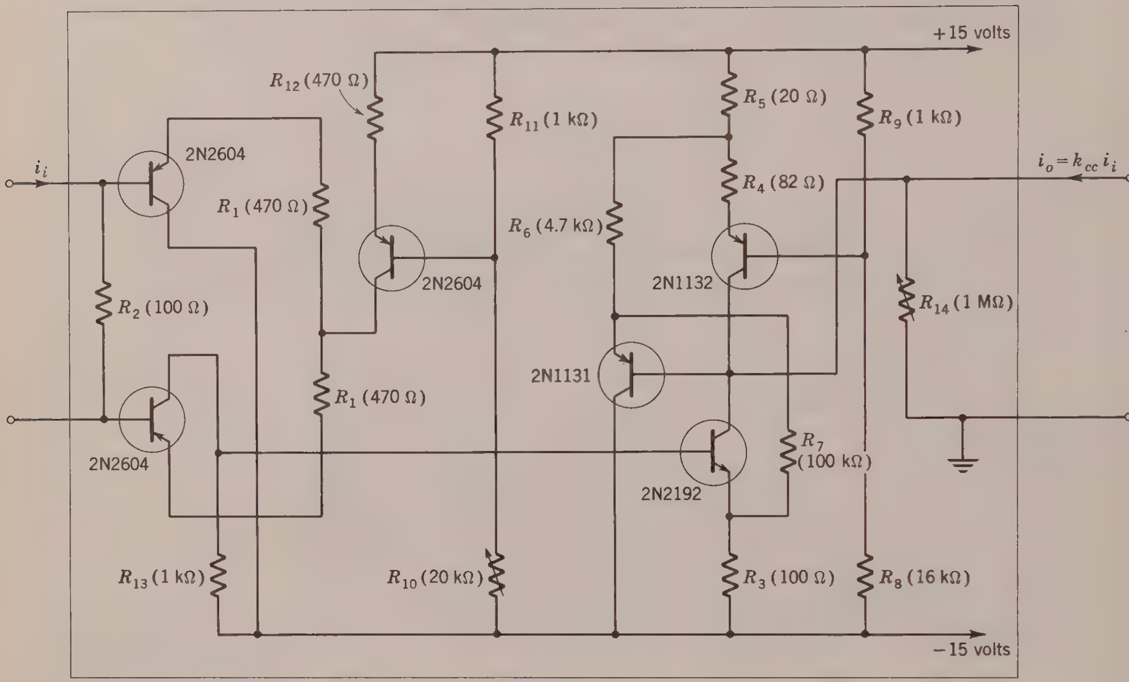


Fig. C-7. Schematic diagram of a current-controlled current source.

shift between the input and output in this case becomes significant above 50 kHz.

Current-controlled voltage source The controlling coefficient k_{cv} in this case is given approximately by

$$k_{cv} = \frac{v_o}{i_i} \approx \left(\frac{R_5 R_{11} R_{15}}{2 R_3 R_4}\right) \left\{\frac{-R_9}{(R_1 + R_{11}) - R_6[(R_1 + R_{11})/(R_6 + R_7)]}\right\} \tag{C-2}$$

If R_6 is very small, a more accurate approximation is obtained by replacing R_{11} in Eq. (C-2) by the parallel combination of R_1 and R_{11}. The qualitative information described above for the voltage-controlled voltage source also applies to this circuit.

Current-controlled current source The controlling coefficient k_{cc} is given approximately by

$$k_{cc} = \frac{i_o}{i_i} \approx \frac{R_2 R_{13}}{2 R_1 R_3} \tag{C-3}$$

The input portion of this circuit (consisting of the three transistors 2N2604 and the resistors R_1, R_2, R_{10}, R_{11}, R_{12}, and R_{13}) constitutes a differential amplifier. When adjusting the gain of this differential amplifier, R_{10} must be readjusted for proper bias current. If the peak-to-peak voltage developed across the input terminals is V_p, the bias current must be at least equal to $V_p/2R_1$ in order to ensure proper operation of the two input transistors. The remaining portion of the circuit on the right is the main current source. In adjusting the gain of this output stage, the ratio R_5/R_4 should be kept constant (in order to maintain positive feedback due to the 2N1131 transistor stage), and R_3 must be set equal to $(R_4 + R_5)$ in order to ensure zero output current whenever the input current is zero. If the gain is increased by adjusting the resistors R_3, R_4, R_5, etc., in the output stage, the value of R_6 and R_7 may have to be reduced in order to ensure an infinite shunt resistance. (An increase in the value of R_6 and R_7 may be necessary if gain is reduced by changing the value of R_3, R_4, and R_5.) An alternate method for adjusting the bias of the 2N1132 transistor consists of adjusting the values of R_8 and R_9 for zero output current when the input current is zero. If this is done, it may not be

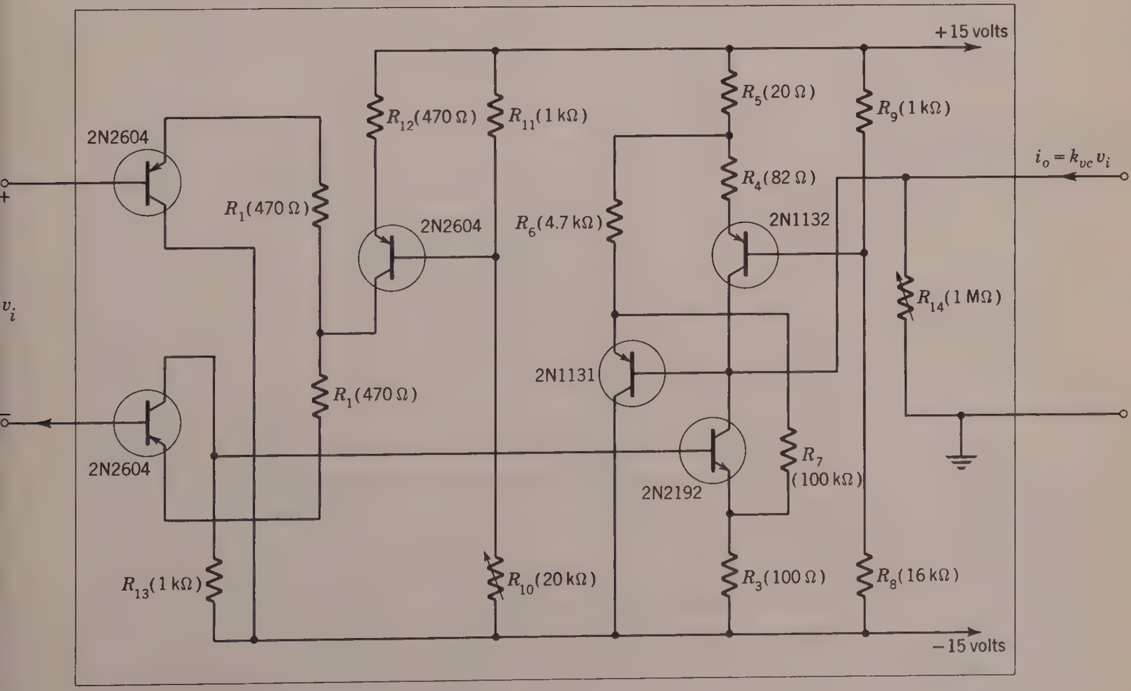


Fig. C-8. Schematic diagram of a voltage-controlled current source.

necessary to change R_4 and R_5 whenever R_3 is changed. In practice, a combination of these methods may be more appropriate.

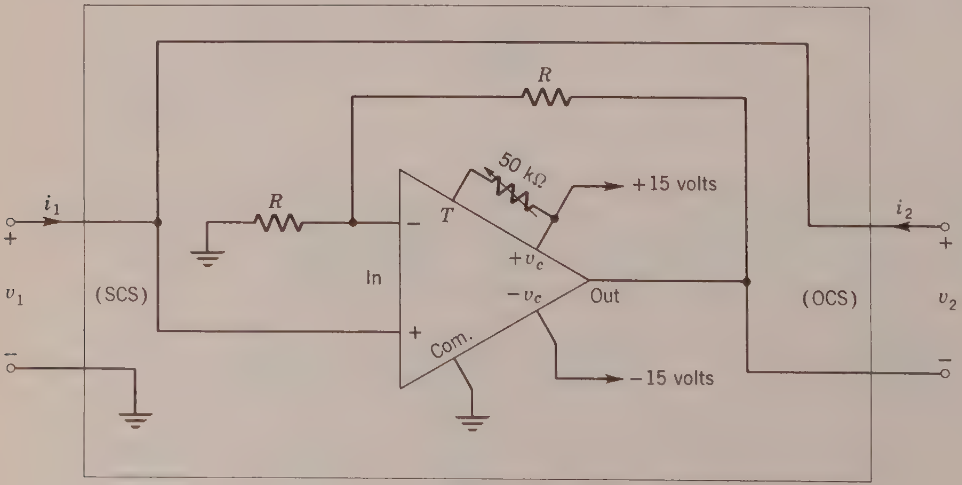
The output current of the 2N1132 transistor is equal approximately to the ratio between the voltage across R_9 (approximately -0.6 volt) and $(R_4 + R_5)$. Similarly, the current input to the collector of the 2N2192 transistor is equal approximately to the ratio between the voltage across R_{13} (approximately -0.6 volt) and R_3. The common-mode limit of the differential input stage is limited by the voltage drop across R_1. If the current bias on this stage is at least $V_p/2R_1$, the voltage drop across either resistor R_1 may be as much as $2V_p$. If the total power supply voltage is 30 volts, the maximum common-mode voltage is about $\pm(15 - 2V_p)$ volts, where, as usual, V_p denotes the peak-to-peak voltage developed across the input terminals.

With the resistor values shown in Fig. C-7, the circuit can deliver a maximum output current of about ±10 ma. The frequency response is flat to about 100 kHz. The output impedance of this circuit can be either positive, negative, or infinite, depending on the adjustment of the 1-MΩ potentiometer across the output terminals.

Voltage-controlled current source The controlling coefficient k_{vc} for this circuit is given approximately by

$$k_{vc} = \frac{R_{13}}{2R_1R_3} \tag{C-4}$$

Fig. C-9. Schematic diagram of a VNIC circuit.



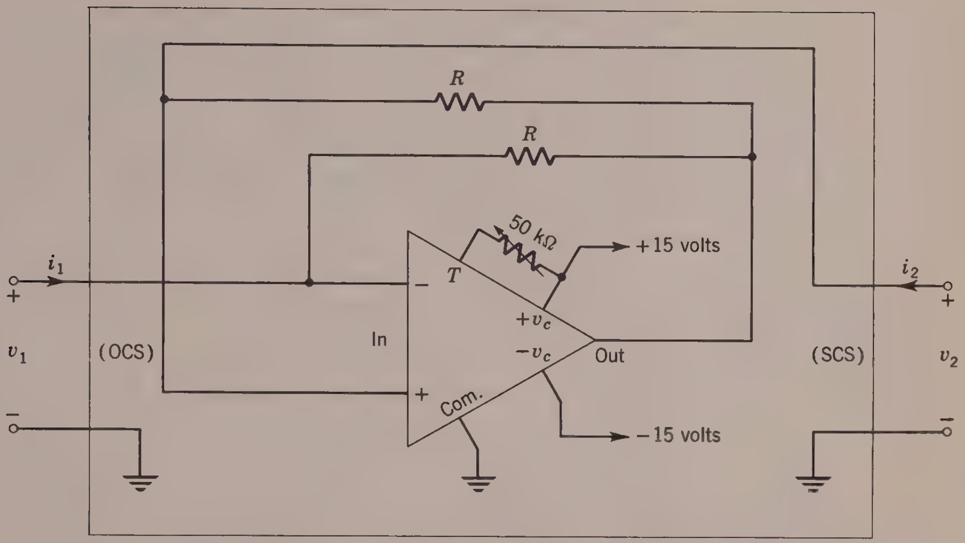


Fig. C-10. Schematic diagram of an INIC circuit.

The qualitative information described above for the current-controlled current source remains valid in this case. The only difference is that whereas V_p in the previous case is equal to the peak-to-peak voltage developed across R_2 (due to the input current), it is equal in this case to the input voltage.

C-3 CIRCUITS FOR SIMULATING NEGATIVE-IMPEDANCE CONVERTERS

The *negative-impedance converter (NIC)* is a linear active two-port network element with the property that whenever an impedance Z is connected across either port 1 or port 2 of the NIC, a negative impedance ($-Z$) is measured across the other port (see Sec. 3-7-3).[1] There are two types of NIC, a VNIC (see Eq. 3-97) and an INIC (see Eq. 3-98). Simple operational amplifier circuits for simulating a VNIC and an INIC are shown in Figs. C-9 and C-10, respectively. The two resistors R may be assigned any convenient value (say, from 10 Ω to 1 MΩ). The circuit shown in Fig. C-10 is preferred in practice because of its common-ground feature.

Because of the presence of parasitic inductances and capacitances, any NIC circuit is potentially unstable. While it is true that either port of the NIC has the same impedance transformation property, it turns out that in order for stable operation to be possible, only one port of the NIC will permit a high-impedance load to be connected across it, while the other port will permit only a low-impedance load. Since in the limit a high-impedance load becomes an open circuit and a low-impedance load becomes a short circuit, the high-impedance port is said to be open-circuit stable (OCS), whereas the low-impedance port is said to be short-circuit stable (SCS). It is common practice, therefore, to label the two ports of an NIC circuit according to whether they are OCS or SCS. For the VNIC circuit shown in Fig. C-9, port 1 is SCS, while port 2 is OCS. The opposite is true for the INIC circuit shown in Fig. C-10.

---

[1] A more complete coverage of negative-impedance converters can be found in L. P. Huelsman, "Theory and Design of Active RC Circuits," chap. 4, McGraw-Hill Book Company, New York, 1968.

C-4 CIRCUITS FOR SIMULATING SCALORS

Two types of scalors are defined in Sec. 3-7-1, namely, a *voltage scalor* and a *current scalor.* A practical circuit for simulating a voltage scalor is shown in Fig. C-11, and that for a current scalor is shown in Fig. C-12. The operational amplifier in Fig. C-11 serves as a voltage-controlled voltage source. The two-transistor symmetrical emitter-follower is used to increase the range of the output current. The voltage scale factor k_v of the voltage scalor is given approximately by

$$k_v \approx 1 + \frac{R_1}{R_2} \tag{C-5}$$

The current scalor circuit in Fig. C-12 exhibits a high degree of symmetry. Hence, if the two transistors are matched, and if the transistors have a high beta, then the following approximate expression can be easily derived for the current scale factor k_i:

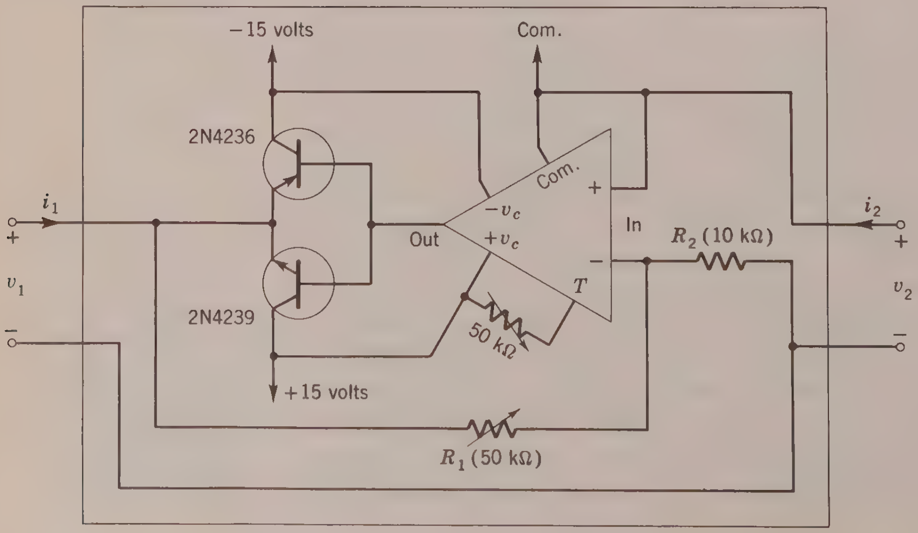


Fig. C-11. Schematic diagram of a voltage scalor.

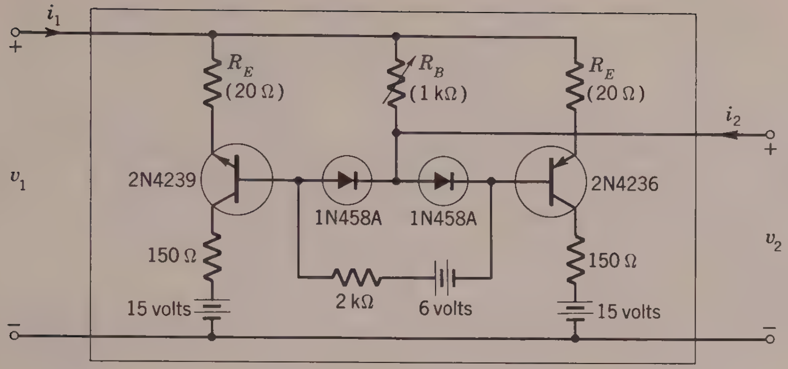


Fig. C-12. Schematic diagram of a current scalor.

$$k_i \approx 1 + \frac{R_B}{R_E} \tag{C-6}$$

Observe that v_2 in this circuit is not exactly equal to v_1 because of the voltage drop across R_B. However, since the current i_2 is generally much smaller than i_1, and since R_B is a relatively small resistance, we have $v_2 \approx v_1$. Finally, observe that the two junction diodes are used only for biasing purposes.

C-5 CIRCUITS FOR SIMULATING ROTATORS

A *rotator* may be simulated in practice by either a π network or a T network of three linear resistors (see Fig. 3-32). If we pick the π network whenever the angle of rotation lies in the range $0° < \theta < 180°$ and the T network whenever the range is $-180° < \theta < 0°$, then only one negative resistor will be needed. This negative resistor can be easily simulated by connecting a positive resistor of the same magnitude across an appropriate port of the VNIC circuit shown in Fig. C-9 or the INIC circuit shown in Fig. C-10. Because of the output current limitation of the operational amplifier, only a low-power rotator can be obtained using these circuits. A high-power INIC circuit designed specifically for this purpose is shown in Fig. C-13. Whether the negative resistance is to be obtained across an OCS port or a SCS port depends on both the angle of rotation θ and the nature of the v-i curve to be rotated.[1]

C-6 CIRCUITS FOR SIMULATING REFLECTORS

The *reflector* is defined in Sec. 3-7-3 for the purpose of reflecting a given v-i curve with respect to some straight line through the

---

[1] L. O. Chua, The Rotator–A New Network Component, *Proc. IEEE*, vol. 55, no. 9, pp. 1566–1577, September, 1967.

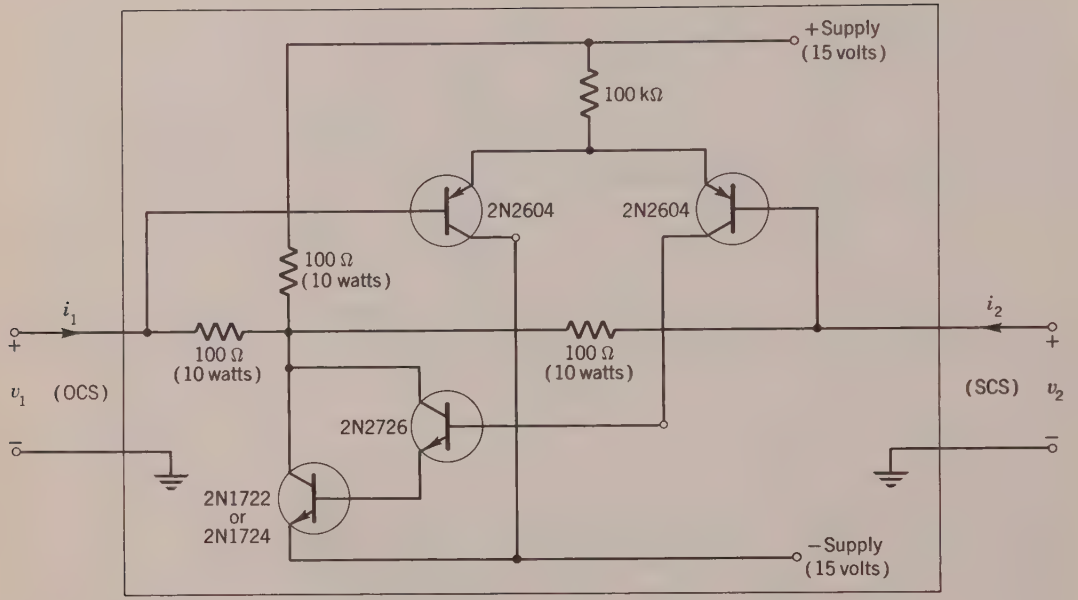


Fig. C-13. Schematic diagram of a high-power INIC circuit.

origin. Two basic realizations of the reflector are given in Fig. 3-34. These circuits can be simulated by the operational amplifier circuits shown in Figs. C-14 and C-15. The resistor values for these circuits were chosen so that the v-i curve will be reflected in the milliampere-volt plane. In order for all resistors in these circuits to be positive, the reflector circuit shown in Fig. C-14 should be used only if the angle of reflection θ lies in the range $0° < \theta < 90°$. Similarly, for the range $90° < \theta < 180°$ the reflector circuit shown in Fig. C-15 should be used. The circuit shown in Fig. C-13 can be used in place of the INIC circuit in Figs. C-14 and C-15.

C-7 CIRCUITS FOR SIMULATING MUTATORS

The *mutator* is defined in Sec. 3-8 for the purpose of transforming one type of network element into another. An R-L mutator will transform a nonlinear resistor into a nonlinear inductor, and vice versa. An R-C mutator will transform a nonlinear resistor into a nonlinear capacitor, and vice versa. A third type of mutator, the C-L mutator, will transform a nonlinear capacitor into a nonlinear inductor, and vice versa. This type of mutator is shown in Sec. 3-8-3 to be a *gyrator* and can, therefore, be simulated by the reflector circuit shown in Fig. C-14 with $\theta = 45°$.

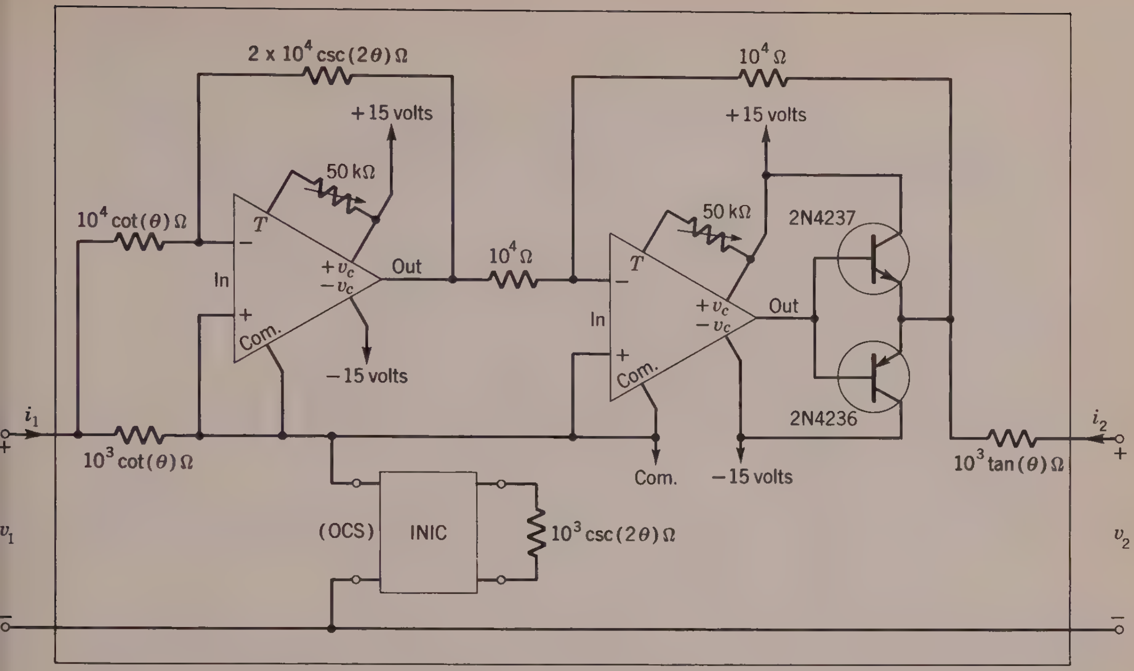
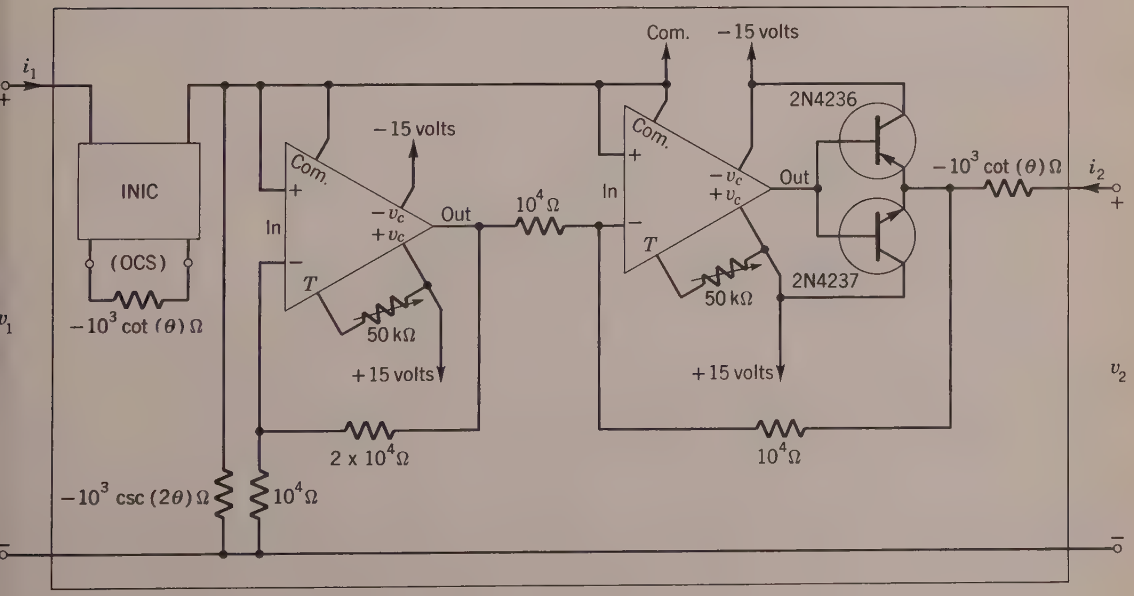


Fig. C-14. Schematic diagram of a reflector circuit for $0° < \theta < 90°$.

Fig. C-15. Schematic diagram of a reflector circuit for $90° < \theta < 180°$.



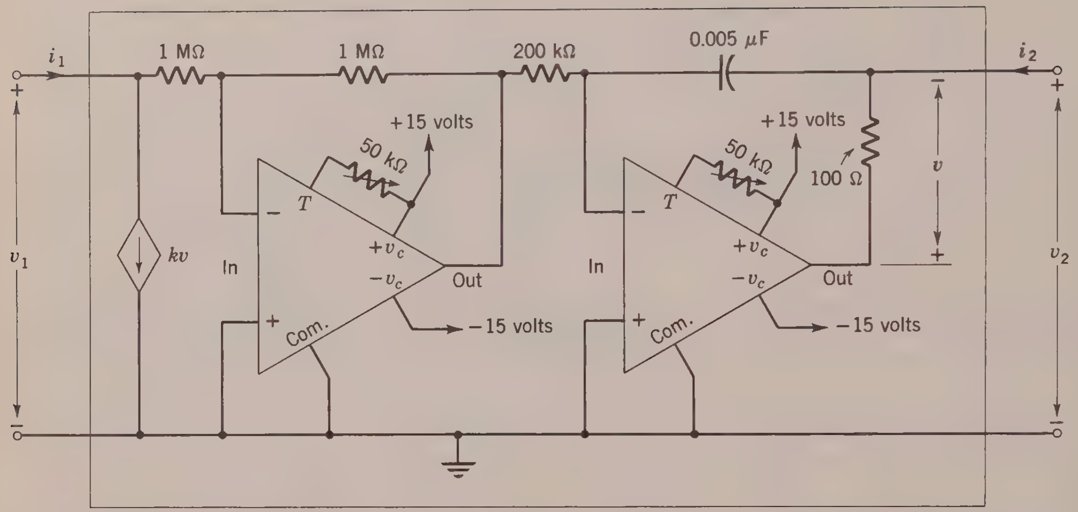
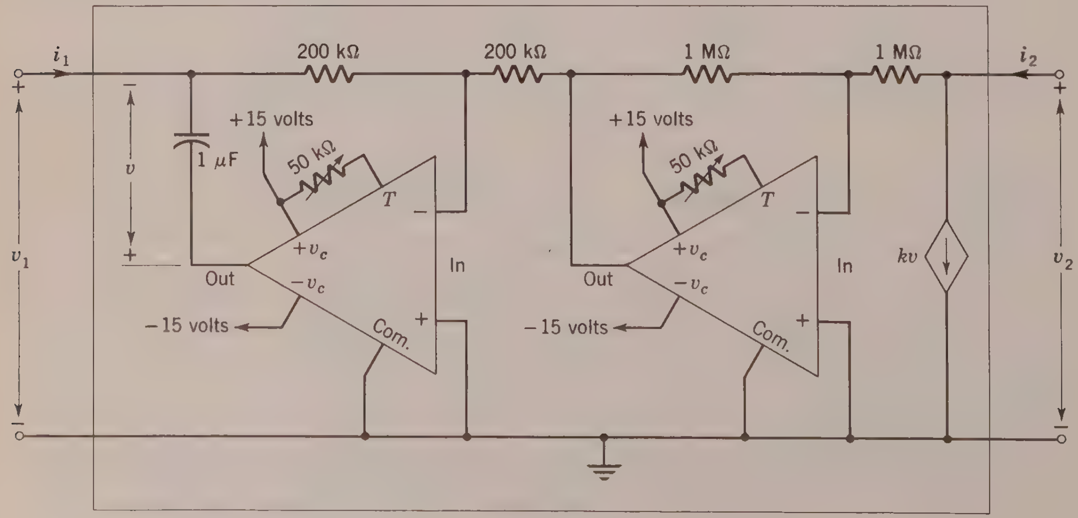


Fig. C-16. Schematic diagram of an *R-L* mutator circuit.

A practical circuit for simulating an *R-L* mutator is shown in Fig. C-16, and one for an *R-C* mutator is shown in Fig. C-17. Observe that a voltage-controlled current source is required in both cases. The circuit shown in Fig. C-8 can be used for this purpose. An inspection of the two mutator circuits in Figs. C-16 and C-17 shows that, except for the location of the linear capacitor, the circuit in Fig. C-17 is a mirror image of the circuit in Fig. C-16. Hence, with the help of a switch, the same circuit can be used to function either as an *R-L* mutator or as an *R-C* mutator.

Fig. C-17. Schematic diagram of an *R-C* mutator circuit.



APPENDIX D CHARACTERISTIC CURVES OF TYPICAL DEVICES

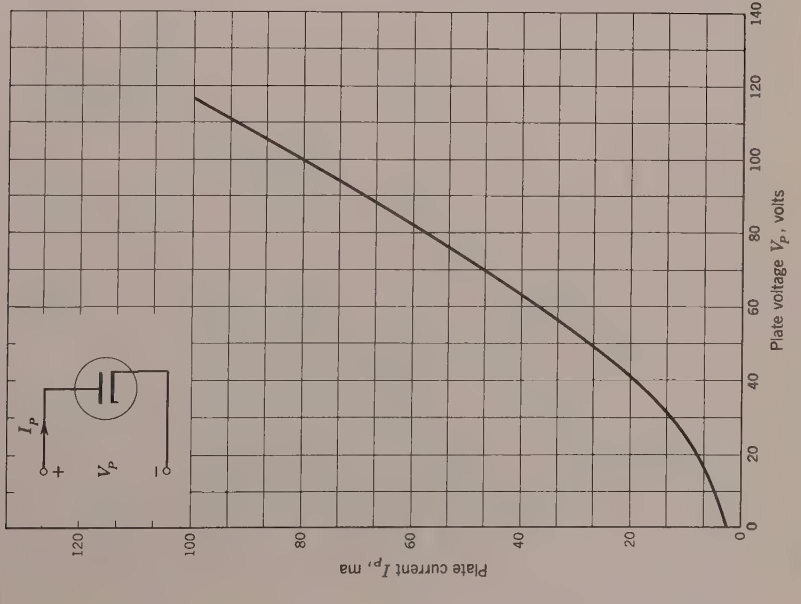


Fig. D-1. Typical characteristics of vacuum diode type 1NXX1.

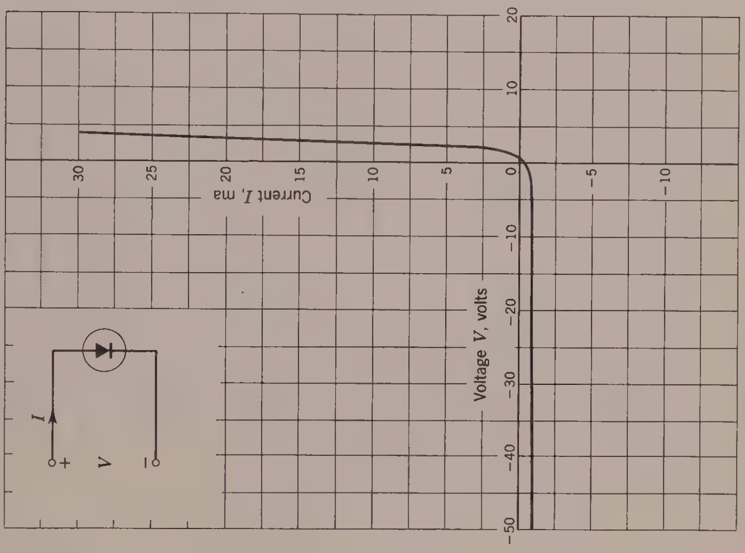


Fig. D-2. Typical characteristics of junction diode type 1NXX2.

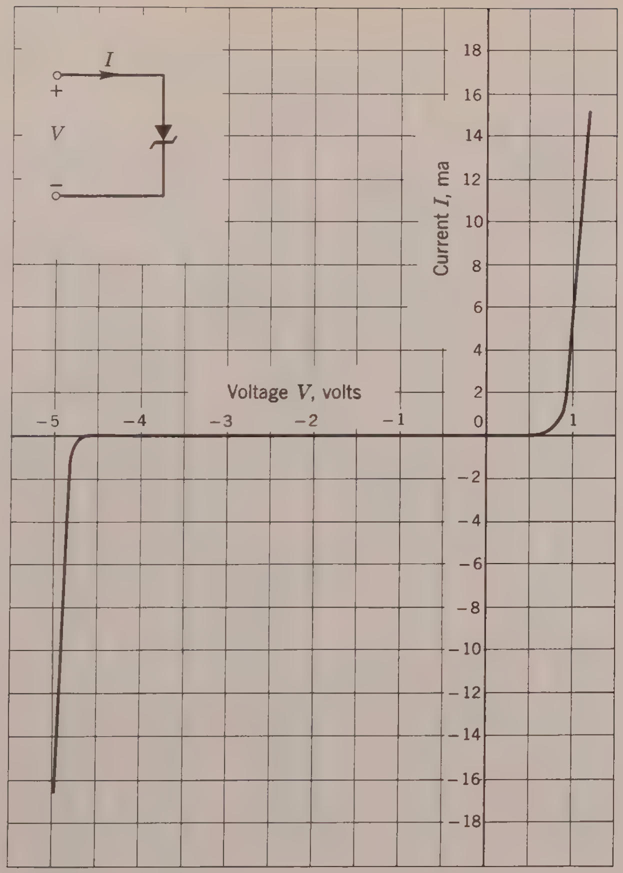


Fig. D-3. Typical characteristics of zener diode type 1NXX3.

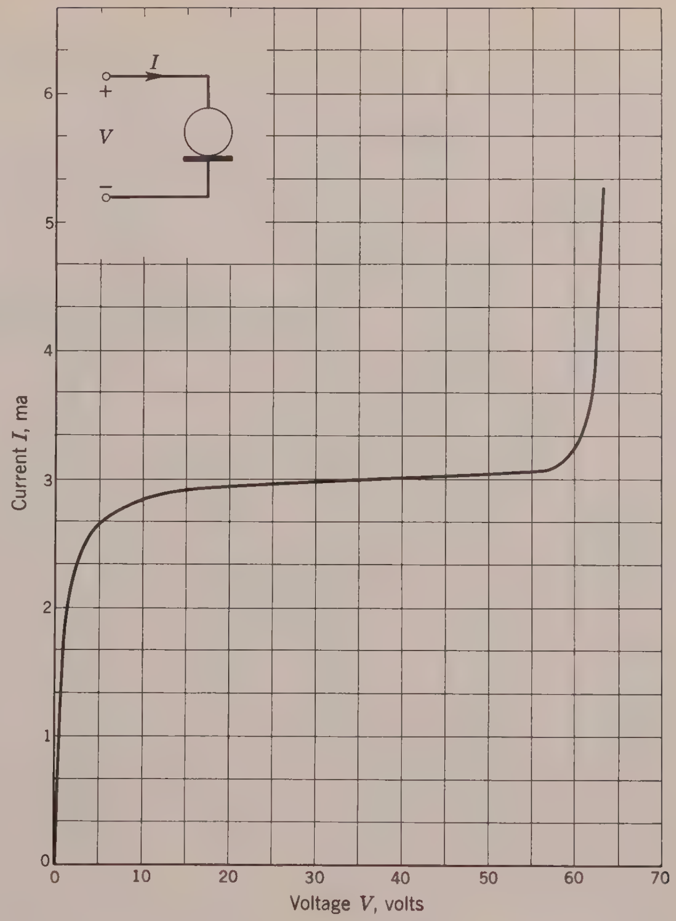


Fig. D-4. Typical characteristics of constant-current diode 1NXX4.

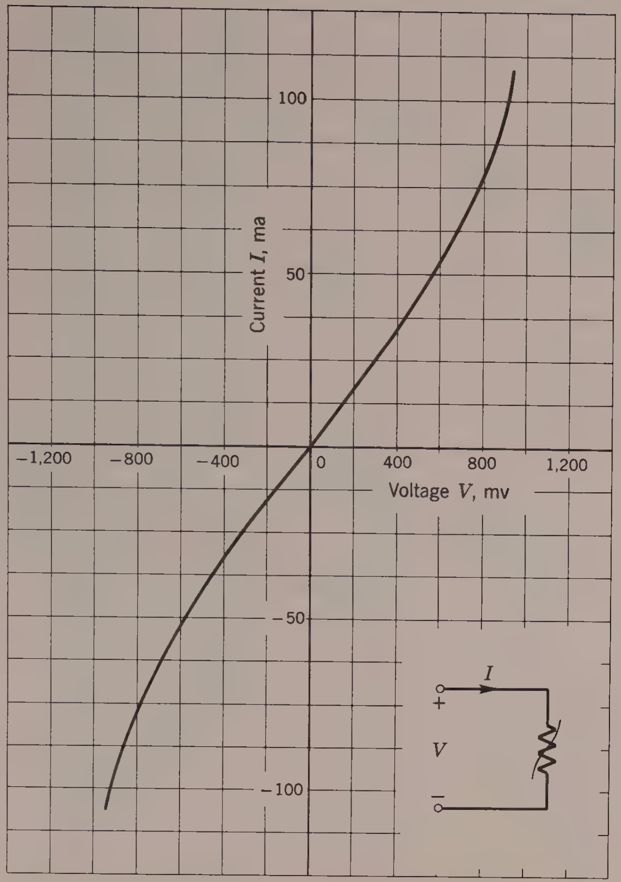


Fig. D-5. Typical characteristics of varistor type 1NXX5.

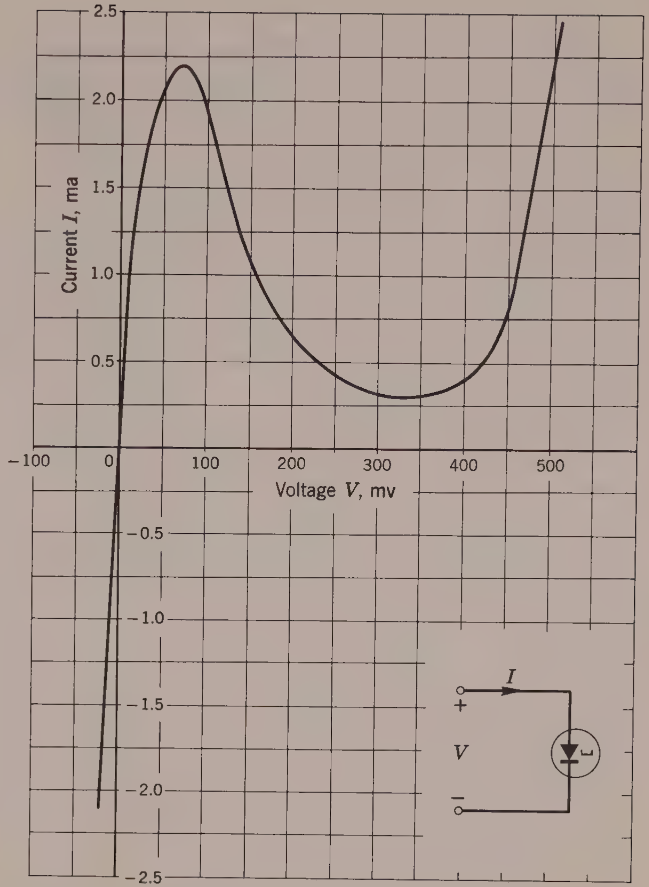


Fig. D-6. Typical characteristics of tunnel diode type 1NXX6.

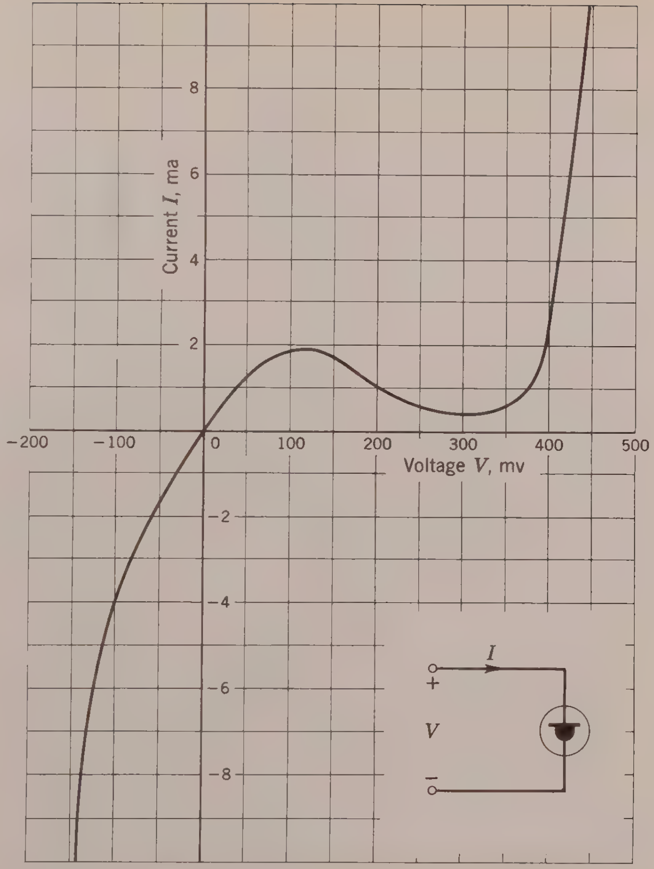


Fig. D-7. Typical characteristics of back diode type 1NXX7.

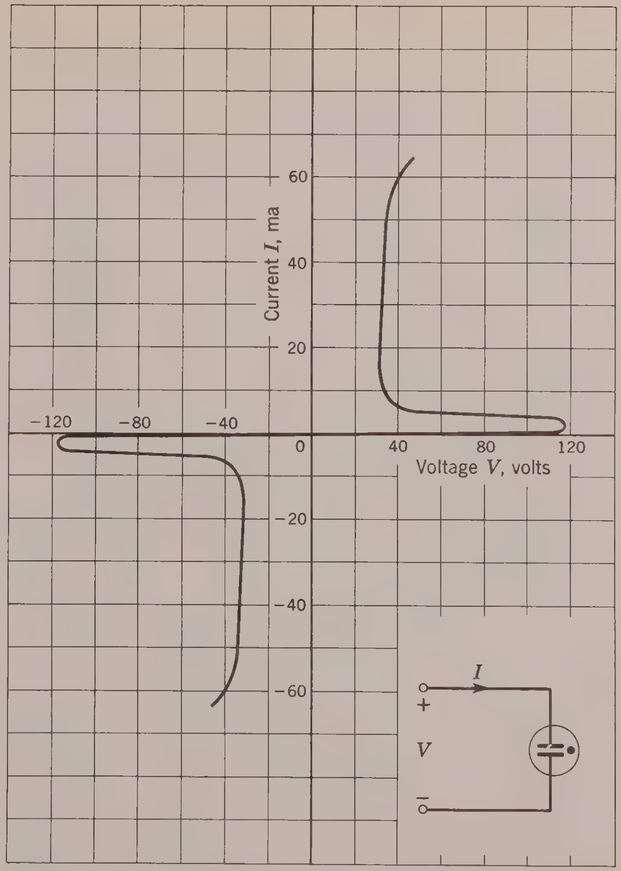


Fig. D-8. Typical characteristics of glow tube type 1NXX8.

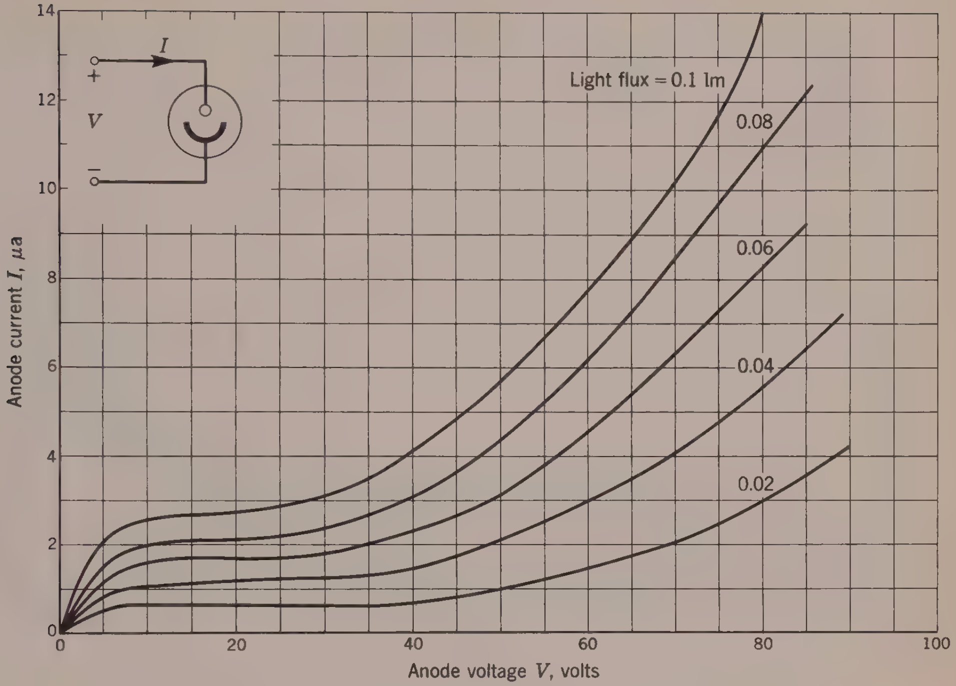
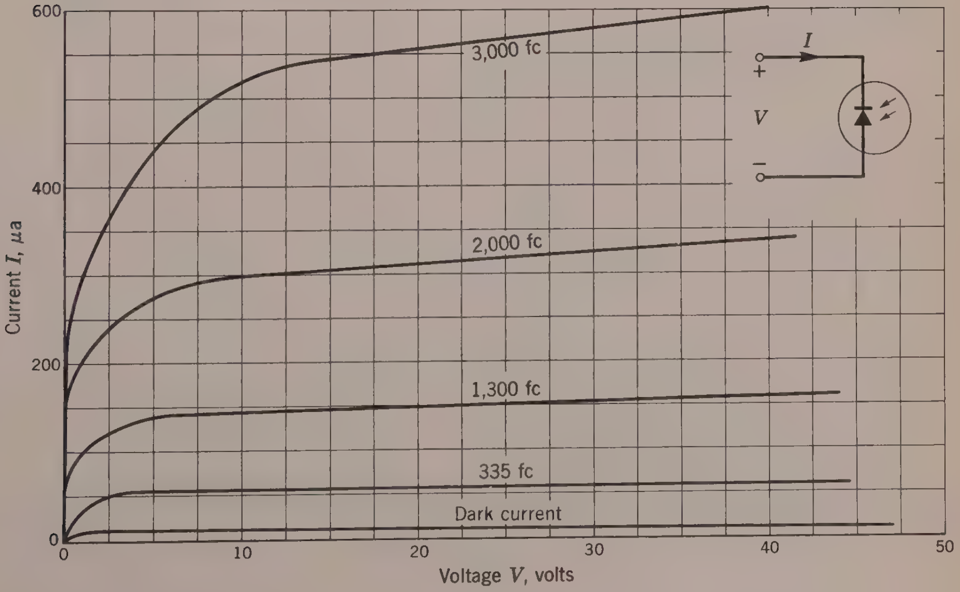


Fig. D-9. Typical characteristics of gas phototube type 1NXX9.

Fig. D-10. Typical characteristics of photodiode type 1NXX10.



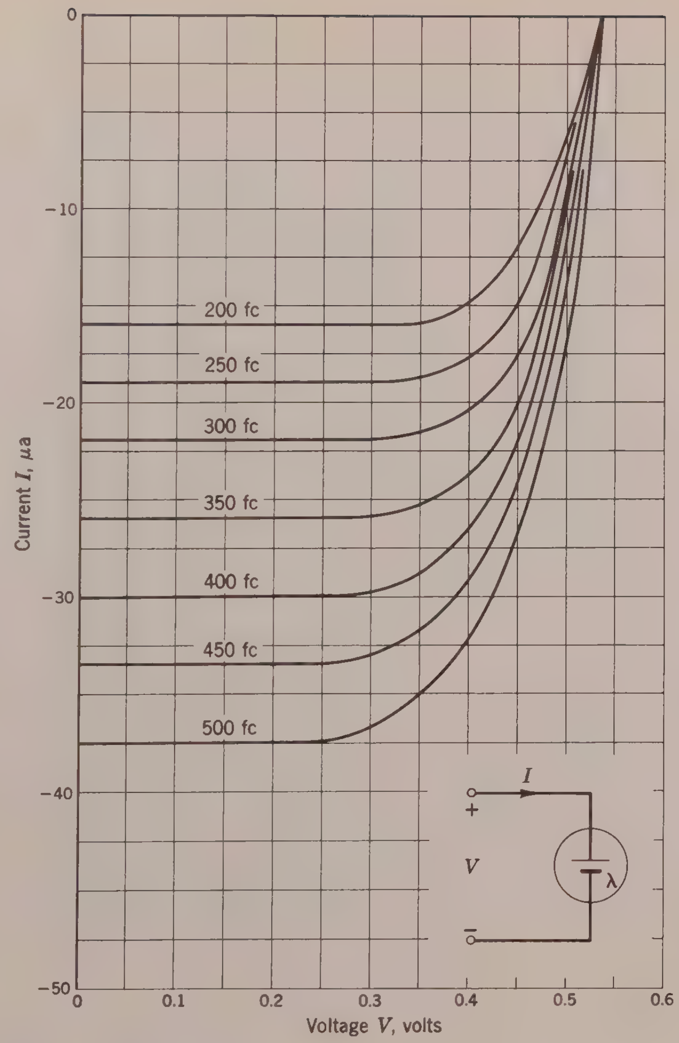


Fig. D-11. Typical characteristics of solar cell type 1NXX11.

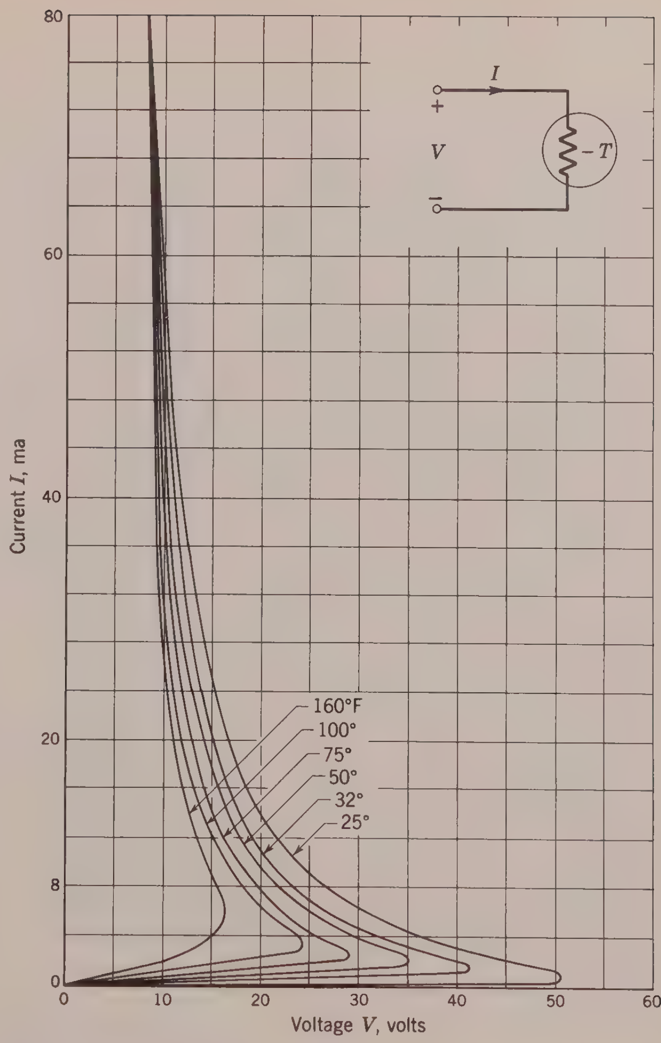


Fig. D-12. Typical characteristics of negative-temperature-coefficient thermistor type 1NXX12

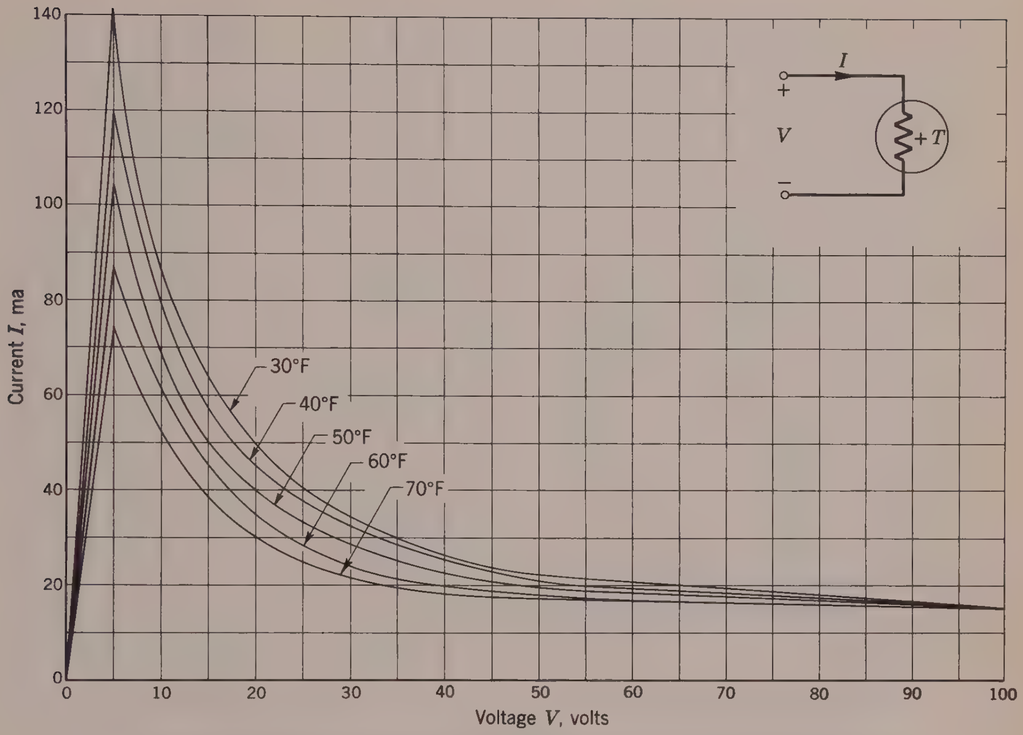
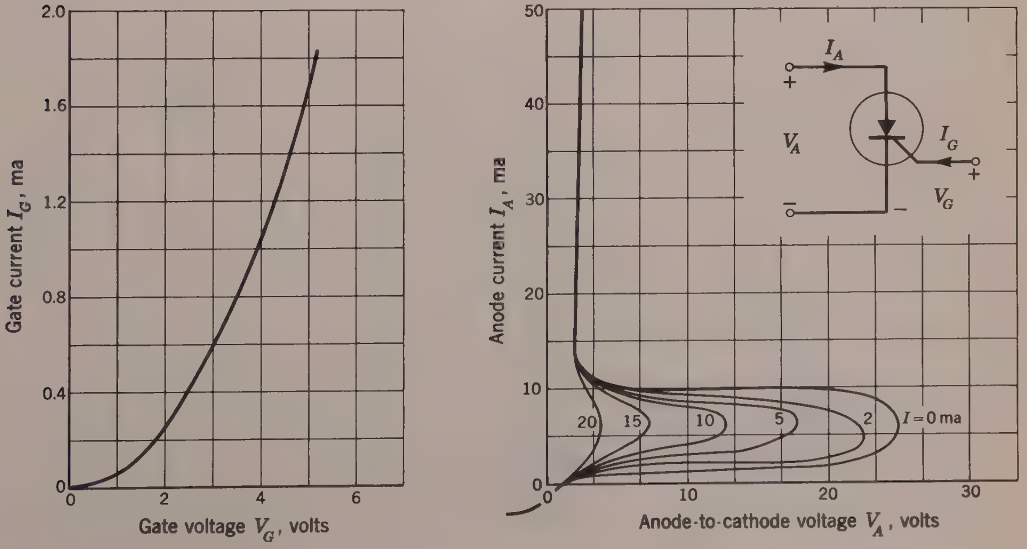


Fig. D-13. Typical characteristics of positive-temperature-coefficient thermistor type 1NXX13.

Fig. D-14. Typical characteristics of SCR type 2NXX14.



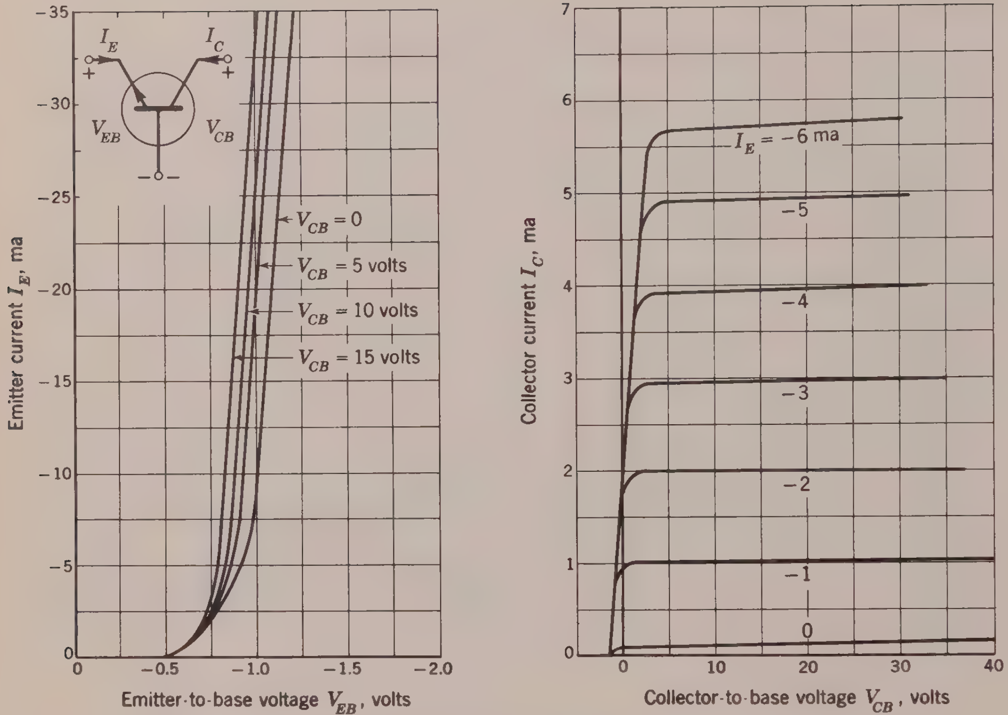
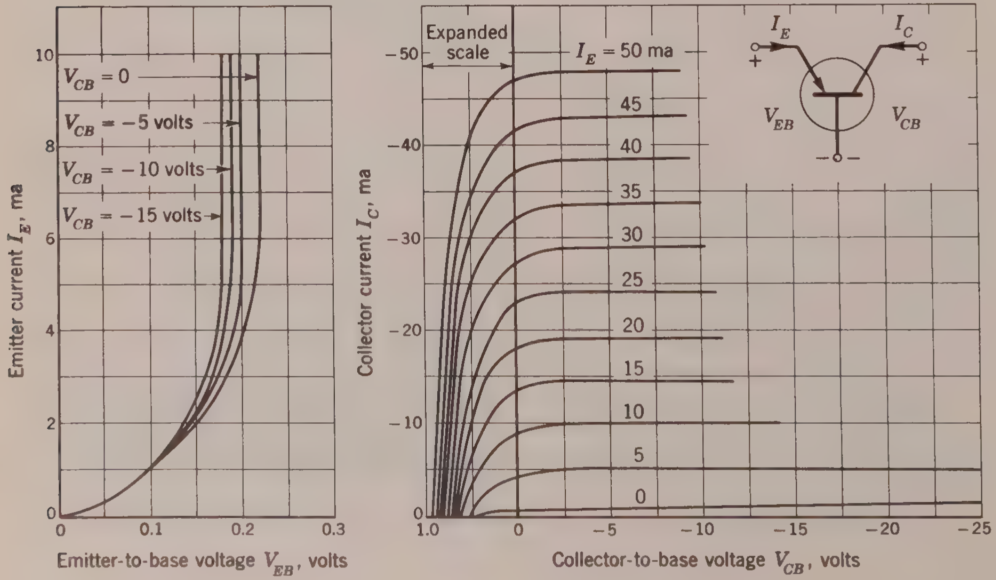


Fig. D-15. Typical characteristics of *n-p-n* transistor type 2NXX15 (common-base configuration).

Fig. D-16. Typical characteristics of *p-n-p* transistor type 2NXX16 (common-base configuration).



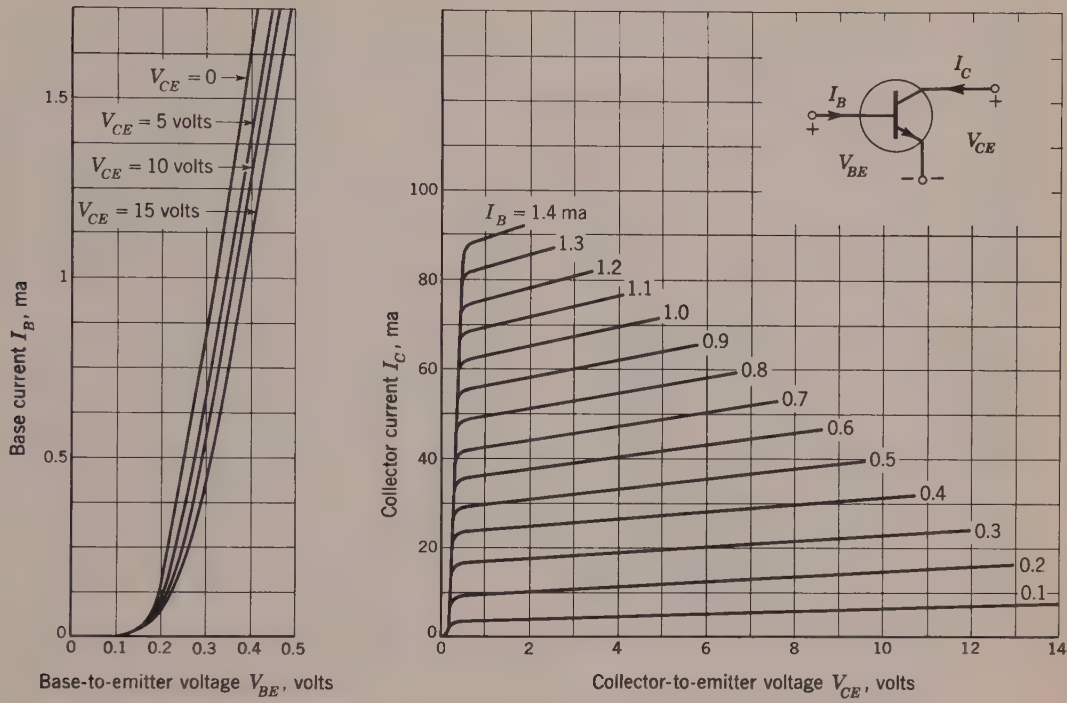
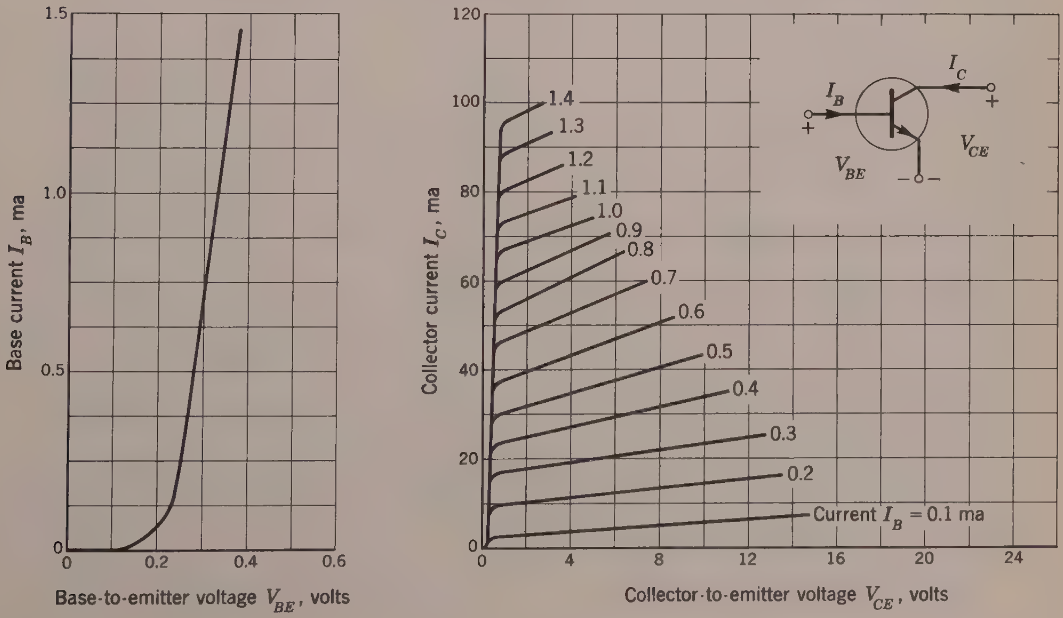


Fig. D-17. Typical characteristics of *n-p-n* transistor type 2NXX17A (common-emitter configuration).

Fig. D-18. Typical characteristics of *n-p-n* transistor type 2NXX17B (common-emitter configuration).



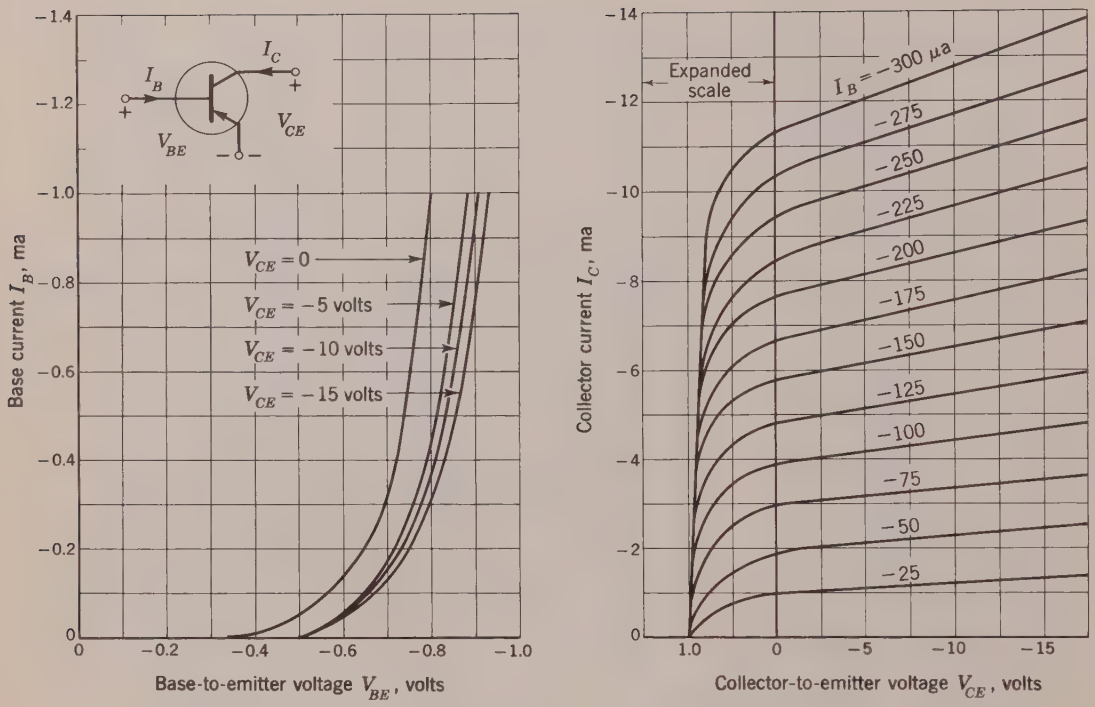
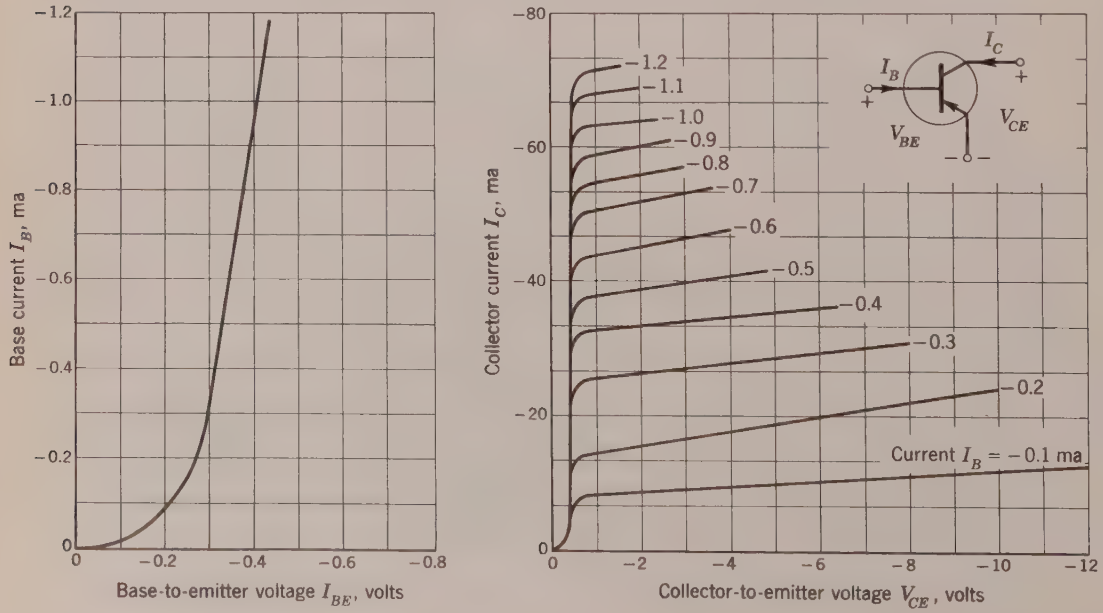


Fig. D-19. Typical characteristics of *p-n-p* transistor type 2NXX18A (common-emitter configuration).

Fig. D-20. Typical characteristics of *p-n-p* transistor type 2NXX18B (common-emitter configuration).



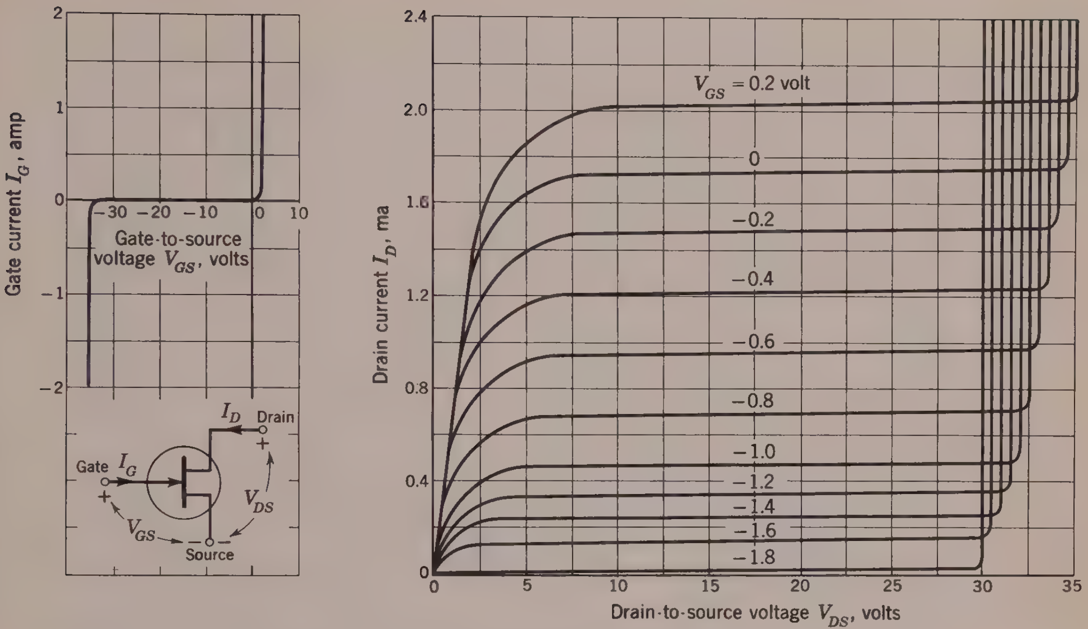
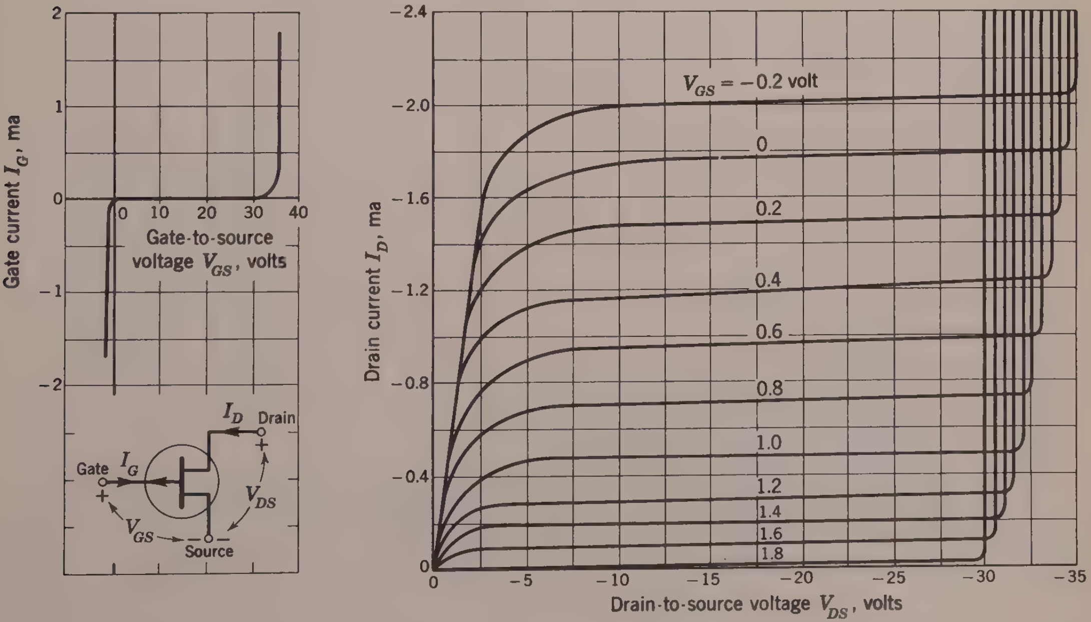


Fig. D-21. Typical characteristics of *n*-channel FET type 2NXX19.

Fig. D-22. Typical characteristics of *p*-channel FET type 2NXX20.



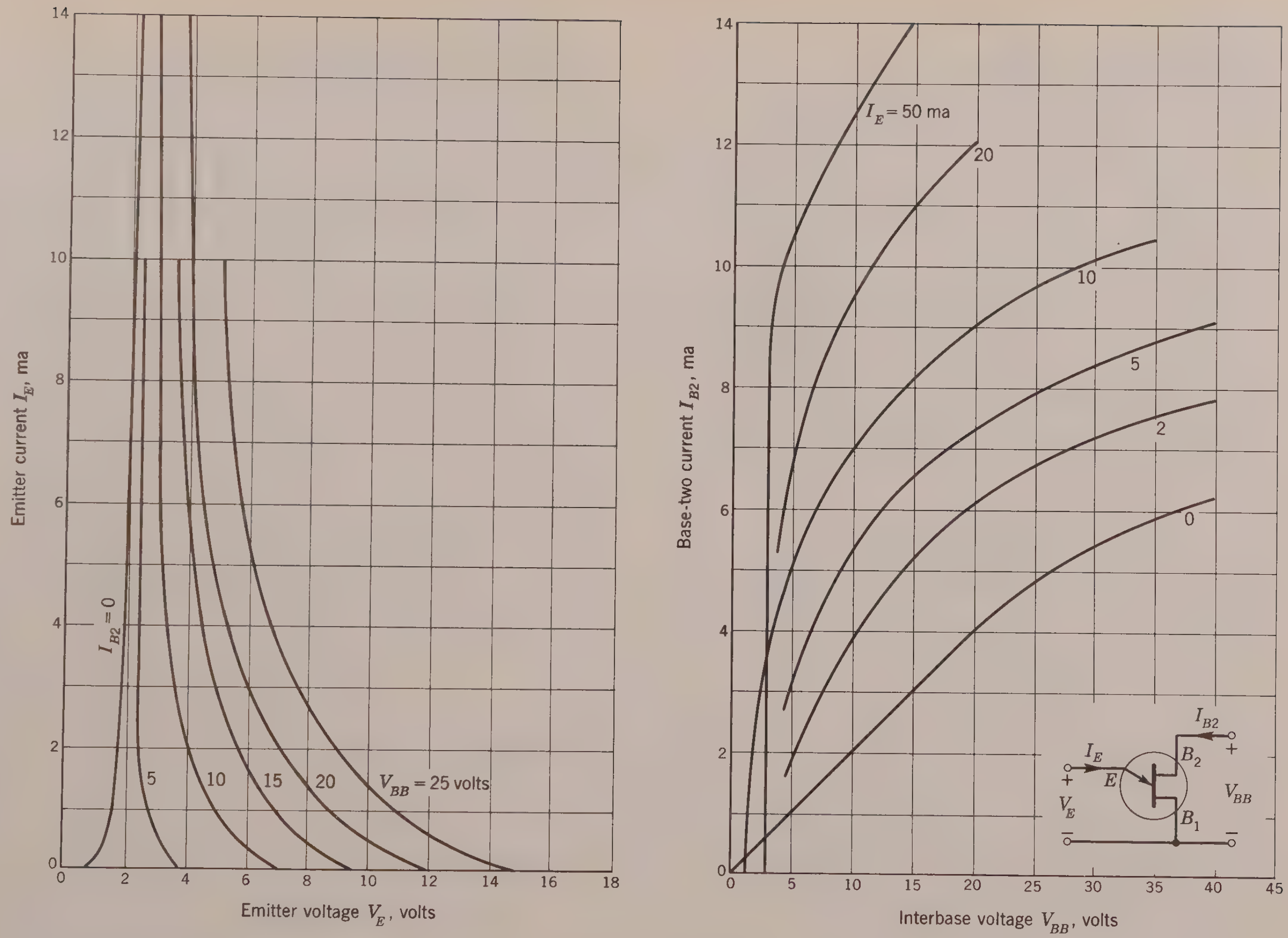


Fig. D-23. Typical characteristics of unijunction transistor type 2NXX21.

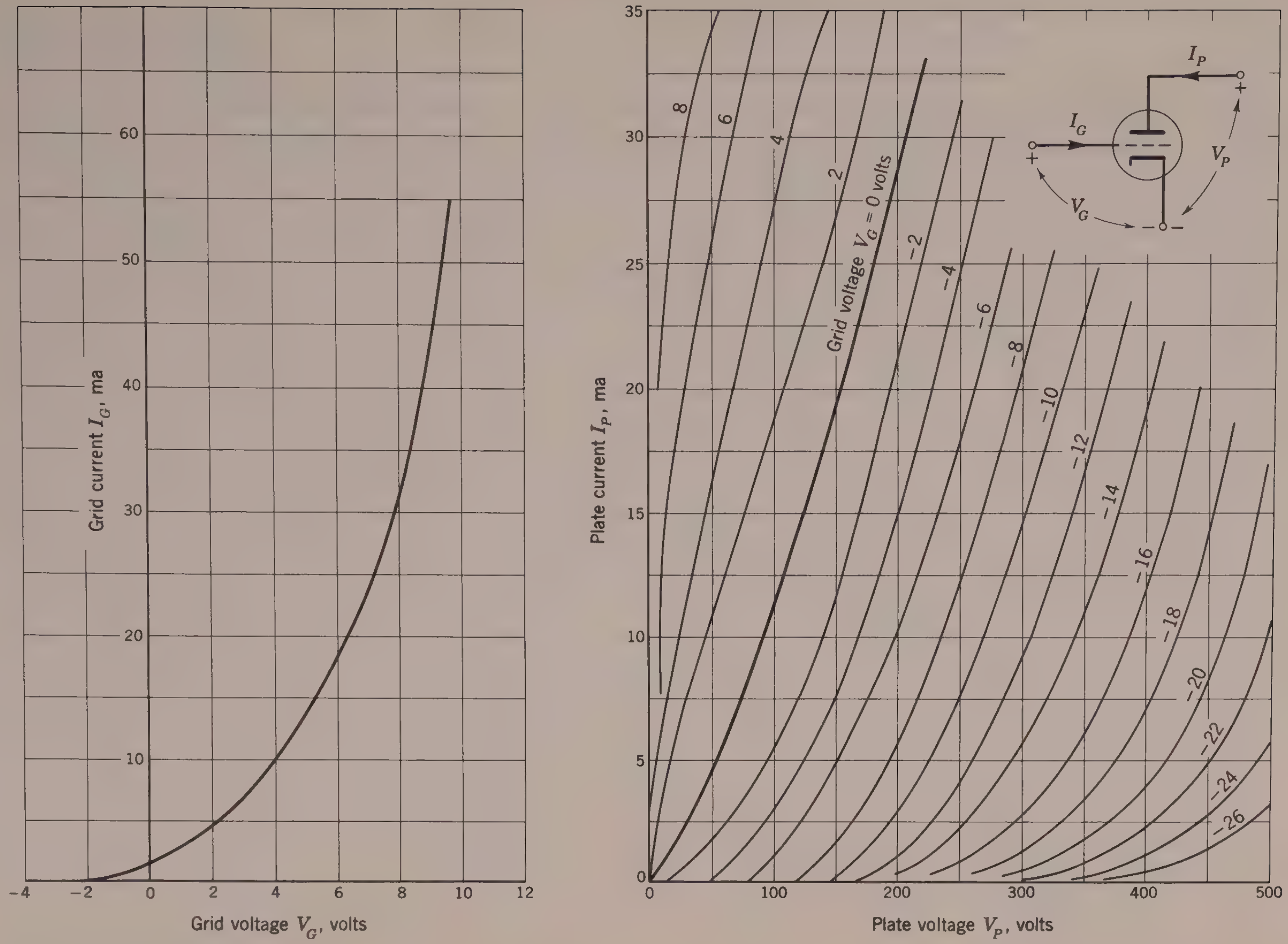


Fig. D-24. Typical characteristics of vacuum triode type 2NXX22.

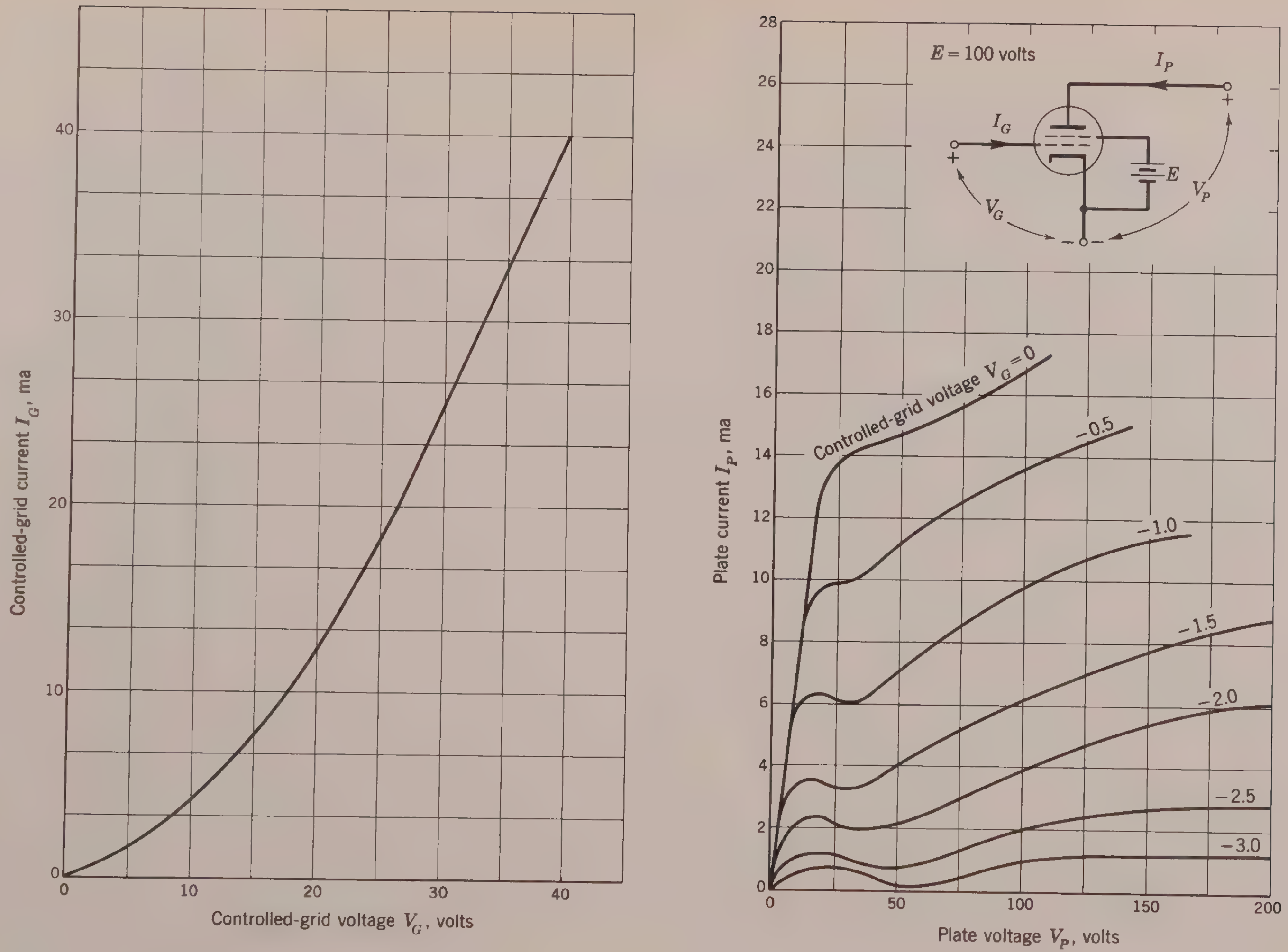


Fig. D-25. Typical characteristics of vacuum tetrode type 2NXX23.

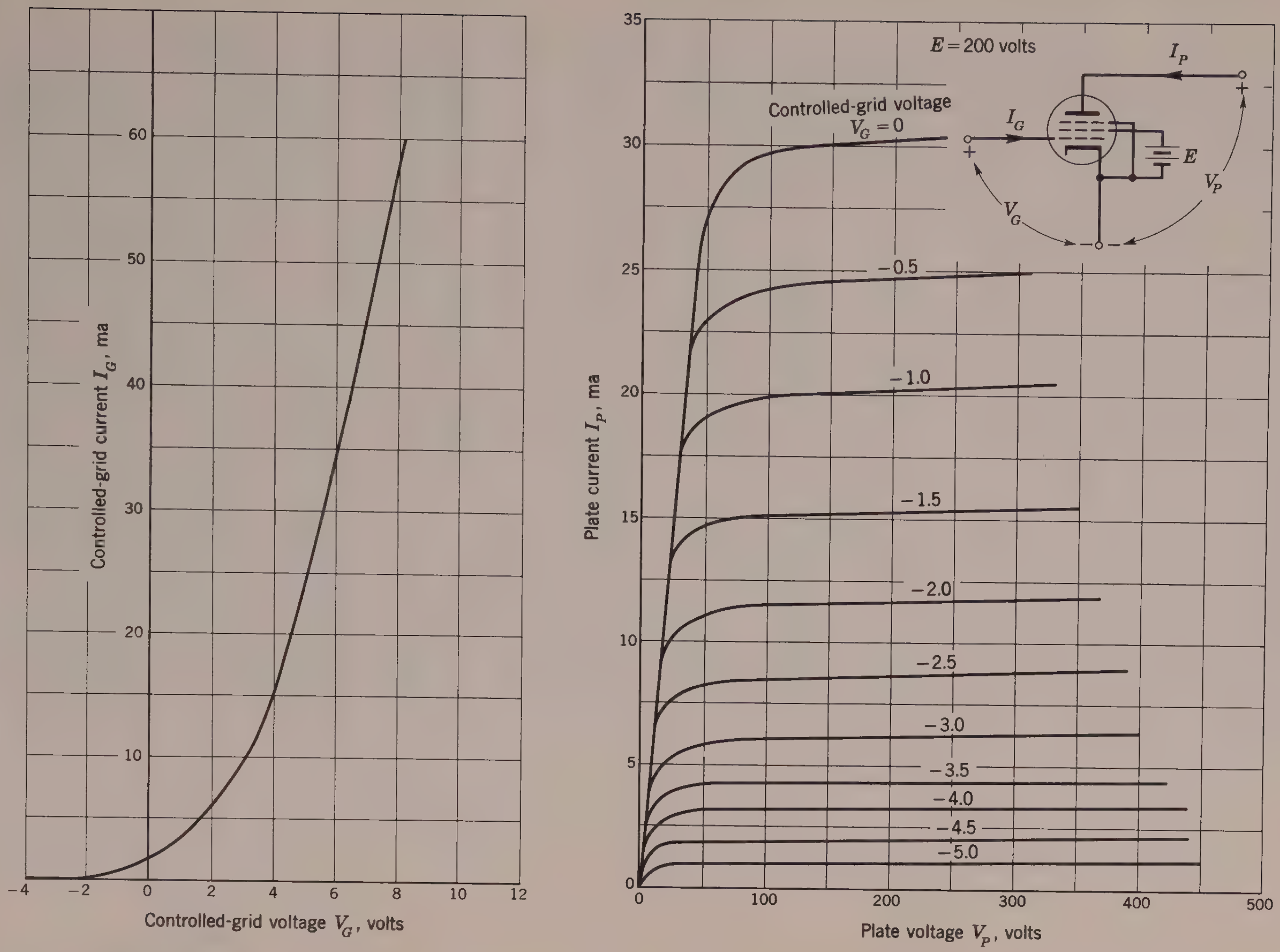


Fig. D-26. Typical characteristics of vacuum pentode type 2NXX24.

APPENDIX E SELECTED LIST OF REFERENCES

E-1 REFERENCES ON BASIC CIRCUIT THEORY

1. Balabanian, Norman: "Fundamentals of Circuit Theory," Allyn and Bacon, Inc., Boston, 1961.
2. Bose, Amar G. and Kenneth N. Stevens: "Introductory Network Theory," Harper and Row, Publishers, Incorporated, New York, 1965.
3. Cruz, Jose B., Jr. and M. E. Van Valkenburg: "Introductory Signals and Circuits," Blaisdell Publishing Company, Waltham, Mass., 1967.
4. Desoer, Charles A. and Ernest S. Kuh: "Basic Circuit Theory," McGraw-Hill Book Company, New York, 1969.
5. Friedland, B., Omar Wing, and R. B. Ash: "Principles of Linear Networks," McGraw-Hill Book Company, New York, 1961.
6. Guillemin, E. A.: "Introductory Circuit Theory," John Wiley & Sons, Inc., New York, 1953.
7. Hayt, W. H., Jr. and J. E. Kemmerly: "Engineering Circuit Analysis," McGraw-Hill Book Company, New York, 1962.
8. Scott, R. E.: "Elements of Linear Circuits," Addison-Wesley Publishing Company, Inc., Reading, Mass., 1965.
9. Skilling, H. H.: "Electrical Engineering Circuits," 2d ed., John Wiley & Sons, Inc., New York, 1965.
10. Van Valkenburg, M. E.: "Network Analysis," 2d ed., Prentice-Hall, Inc., Englewood Cliffs, N.J., 1964.

E-2 REFERENCES ON BASIC ELECTRONIC CIRCUITS AND DEVICES

1. Angelo, E. J.: "Electronic Circuits," 2d ed., McGraw-Hill Book Company, New York, 1963.
2. Gibbons, James F.: "Semiconductor Electronics," McGraw-Hill Book Company, New York, 1966.
3. Gray, Truman S.: "Applied Electronics," John Wiley & Sons, Inc., New York, 1954.
4. Hayt, William H., Jr. and George W. Hughes: "Introduction to Electrical Engineering," McGraw-Hill Book Company, New York, 1968.
5. Millman, Jacob and Christos C. Halkias: "Electronic Devices and Circuits," McGraw-Hill Book Company, New York, 1967.
6. Pederson, Donald O., Jack J. Studer, and John R. Whinnery: "Introduction to Electronic Systems, Circuits, and Devices," McGraw-Hill Book Company, New York, 1966.
7. Searle, C. L., A. R. Boothroyd, E. J. Angelo, Jr., P. E. Gray, and D. O. Pederson: "Elementary Circuit Properties of Transistors," John Wiley & Sons, Inc., New York, 1964.
8. Zimmerman, H. J. and S. J. Mason: "Electronic Circuit Theory," John Wiley & Sons, Inc., New York, 1959.

E-3 REFERENCES ON LINEAR ACTIVE NETWORKS

1. DePian, Louis: "Linear Active Network Theory," Prentice-Hall, Inc., Englewood Cliffs, N.J., 1962.
2. Ghausi, Mohammed Shuaib: "Principles and Design of Linear Active Circuits," McGraw-Hill Book Company, New York, 1965.
3. Huelsman, Lawrence P.: "Circuits, Matrices, and Linear Vector Spaces," McGraw-Hill Book Company, 1963.
4. Huelsman, Lawrence P.: "Theory and Design of Active RC Circuits," McGraw-Hill Book Company, New York, 1968.
5. Linvill, John G. and James F. Gibbons: "Transistors and Active Circuits," McGraw-Hill Book Company, New York, 1961.

6. Mitra, Sanjit Kumar: "Analysis and Synthesis of Linear Active Networks," John Wiley & Sons, Inc., New York, 1969.
7. Su, Kendall L.: "Active Network Synthesis," McGraw-Hill Book Company, New York, 1965.

E-4 REFERENCES ON NONLINEAR ELECTRONIC CIRCUITS

1. Anner, George E.: "Elementary Nonlinear Electronic Circuits," Prentice-Hall, Inc., Englewood Cliffs, N.J., 1967.
2. Chance, B., et al.: "Waveforms," M.I.T. Radiation Laboratory Series, vol. 19, McGraw-Hill Book Company, New York, 1949.
3. Edson, William A.: "Vacuum Tube Oscillators," John Wiley & Sons, Inc., New York, 1953.
4. Glasoe, G. N. and J. V. Lebacqz: "Pulse Generators," M.I.T. Radiation Laboratory Series, vol. 5, McGraw-Hill Book Company, New York, 1948.
5. Groszkowski, J.: "Frequency of Self-oscillations," Pergamon Press, New York, 1964.
6. Joyce, Maurice V. and Kenneth K. Clarke: "Transistor Circuit Analysis," Addison-Wesley Publishing Company, Inc., Reading, Mass., 1961.
7. Littauer, Raphael: "Pulse Electronics," McGraw-Hill Book Company, New York, 1965.
8. Millman, Jacob and Herbert Taub: "Pulse, Digital, and Switching Waveforms," McGraw-Hill Book Company, New York, 1965.
9. Preisman, A.: "Graphical Construction for Vacuum Tube Circuits," McGraw-Hill Book Company, New York, 1943.
10. Reich, Herbert J.: "Functional Circuits and Oscillators," D. Van Nostrand Company, Inc., Princeton, N.J., 1961.
11. Strauss, Leonard: "Wave Generation and Shaping," McGraw-Hill Book Company, New York, 1960.

E-5 REFERENCES ON NONLINEAR NETWORK ELEMENTS

1. Amperex: *Nonlinear Resistors*, Amperex Electronic Corporation, Hicksville, N.Y.

2. Anderson, J. R.: Ferroelectric Materials as Storage Elements for Digital Computers and Switching Systems, *Elec. Eng.*, vol. 71, pp. 916–922, October, 1952.
3. Bakanowski, A. E., N. G. Cranna, and A. Uhlir: Diffused Silicon Nonlinear Cacapitors, *IRE Trans. Electron Devices,* pp. 384–390, October, 1959.
4. Biondi, F. J. (ed.): "Transistor Technology," vol. 3, chap. 10, Device Measurement, pp. 325–373, D. Van Nostrand Company, Inc., Princeton, N.J., 1958.
5. Blackburn, J. F. (ed.): "Components Handbook," M.I.T. Radiation Laboratory Series, vol. 17, McGraw-Hill Book Company, New York, 1948.
6. Bowen, Clarence E.: Forward with Backward Diodes, *Electron. Design,* pp. 44–49, April 26, 1966.
7. Bridgers, H. E., J. H. Scaffs and J. N. Shive: "Transistor Technology," vol. 1, part 4, Principles of Transistor Performance Characterization, pp. 451–604, D. Van Nostrand Company, Inc., Princeton, N.J., 1958.
8. Brownlee, T.: The Calculation of Circuits Containing Thyrite, *General Electric Review,* vol. 37, no. 4, pp. 175–179, April, 1934.
9. Buck, D. A.: The Cryotron–A Superconductive Computer Component, *Proc. IRE,* vol. 44, pp. 482–493, 1956.
10. Chua, Leon O.: The Rotator–A New Network Component, *Proc. IEEE,* vol. 55, no. 9, pp. 1566–1577, September, 1967.
11. Chua, Leon O.: Synthesis of New Nonlinear Network Elements, *Proc. IEEE,* vol. 56, no. 8, pp. 1325–1340, August, 1968.
12. Gentry, F. E., F. W. Gutzwiller, Nick Holonyak, Jr., and E. E. Von Zastrow: "Semiconductor Controlled Rectifiers," Prentice-Hall, Inc., Englewood Cliffs, N.J., 1964.
13. Geyger, William A.: "Nonlinear Magnetic Control Devices," McGraw-Hill Book Company, New York, 1964.
14. Gibbons, James F.: "Semiconductor Electronics," McGraw-Hill Book Company, New York, 1966.
15. Hall, R. N.: Tunnel Diodes, *IRE Trans. Electron Devices,* vol. ED-7, no. 1, pp. 1–9, January, 1960.

16. Hollmann, Hans E.: Semiconductive Colloidal Suspensions with Nonlinear Properties, *J. Appl. Phys.,* vol. 21, pp. 402–413, May, 1950.
17. Hollmann, Hans E.: Nonlinear Elements and Some Applications in AF and RF Circuits, *Proc. Nat. Electron. Conf.,* vol. 7, pp. 130–140, 1951.
18. Howson, D. P.: Nonlinear-Capacitance Devices, *Electron. Letters,* vol. 1, no. 2, pp. 49–50, April, 1965.
19. Jaffe, Bernard: Anti-ferroelectric Ceramics with Field-enforced Transitions: A New Nonlinear Circuit Element, *Proc. IRE,* pp. 1264–1267, August, 1961.
20. Kallman, H. E.: The Assistor, A Component with Bipolar Negative Resistance, *Proc. IRE,* vol. 50, pp. 2138–2139, October, 1962.
21. Lane, R. N. and D. B. Cameron: Current Integration with Solion Liquid Diodes, *Electronics,* pp. 53–55, Feb. 27, 1959.
22. Lesk, I. A.: Germanium p-n-p-n Switches, *IRE Trans. Electron Devices,* vol. ED-6, no. 1, pp. 28–35, January, 1959.
23. Loebner, E. E.: Opto-electronic Devices and Networks, *Proc. IRE,* pp. 1897–1906, December, 1955.
24. Massachusetts Institute of Technology, Electrical Engineering Staff: "Magnetic Circuits and Transformers," John Wiley & Sons, Inc., New York, 1943.
25. Meyer, H. J. G., G. Ahsmann, J. W. van der Laarse, and E. H. A. M. van der Wee: A New Class of Low Pressure Arc Columns with Positive V-I Characteristics, *Philips Research Reports,* vol. 22, no. 3, pp. 209–218, June, 1967.
26. Mueller, C. W. and J. Hilibrand: The "Thyristor"–A New High-speed Switching Transistor, *IRE Trans. Electron Devices,* vol. ED-5, pp. 2–5, January, 1958.
27. Oxley, T. H.: Backward Diodes as Mixers at Microwave Frequencies, *J. Electron. Control,* vol. 17, pp. 1–17, July, 1964.
28. Prywes, N. S. (ed.): "Amplifier and Memory Devices," McGraw-Hill Book Company, New York, 1965.
29. Saburi, O. and K. Wakino: Processing Techniques and Applications of Positive Temperature Coefficient Thermistors, *IEEE Trans. Component Pts,* pp. 53–67, June, 1963.

30. Scarr, R. W. A. and R. A. Setterington: Thermistors, Their Theory, Manufacture, and Application, *IRE Trans. Component Pts.,* pp. 6–22, March, 1961.

31. Schwertz, F. A. and J. J. Mazenko: Nonlinear Semiconductor Resistors, *J. Appl. Phys.,* vol. 24, no. 8, pp. 1015–1024, August, 1953.

32. Sevin, Leonce J.: "Field-effect Transistors," McGraw-Hill Book Company, New York, 1965.

33. Statz, H. and R. A. Pucel: The Spacistor, A New Class of High-frequency Semiconductor Device, *Proc. IRE,* pp. 317–324, March, 1957.

34. Stuetzer, O. M.: Junction Fieldistors, *Proc. IRE,* pp. 1377–1381, November, 1952.

35. Summer, W.: "Photosensitors," Chapman & Hall, Ltd., London, 1957.

36. Suran, J. J.: Low Frequency Circuit Theory of the Double-base Diode, *IRE Trans. Electron Devices,* vol. ED-2, no. 2, pp. 40–48, April, 1955.

37. Torrey, Henry C. and Charles A. Whitmer: "Crystal Rectifiers," M.I.T. Radiation Laboratory Series, vol. 15, McGraw-Hill Book Company, New York, 1948.

38. Wallmark, J. T., L. Varettoni, and H. Ur: The Tunnel Resistor, *IEEE Trans. Electron Devices,* vol. ED-10, no. 6, pp. 359–363, November, 1963.

39. Wallmark, J. Torkel and Harwick Johnson: "Field Effect Transistors," Prentice-Hall, Inc., Englewood Cliffs, N.J., 1966.

E-6 REFERENCES ON INTEGRATED CIRCUITS

1. Burger, R. M. and R. P. Dunovan: "Fundamentals of Silicon Integrated Device Technology," vol. 1 & 2, Prentice-Hall, Inc., Englewood Cliffs, N.J., 1967.

2. Chirlian, Paul M.: "Integrated and Active Network Analysis and Synthesis," Prentice-Hall, Inc., Englewood Cliffs, N.J., 1967.

3. Lin, H. C.: "Integrated Electronics," Holden-Day, San Francisco, Calif., 1967.

4. Meyer, Charles S., David K. Lynn, and Douglas J. Hamilton: "Analysis and Design of Integrated Circuits," McGraw-Hill Book Company, New York, 1967.
5. Newcomb, Robert W.: "Active Integrated Circuit Synthesis," Prentice-Hall, Inc., Englewood Cliffs, N.J., 1968.
6. Schwartz, Seymour: "Integrated Circuit Technology," McGraw-Hill Book Company, New York, 1967.

E-7 REFERENCES ON DEVICE MODELING

1. Andronov, A. A., A. A. Vitt, and S. E. Khaikin: "Theory of Oscillators," Addison-Wesley Publishing Company, Inc., Reading, Mass., 1966.
2. Gray, P. E., D. DeWitt, A. R. Boothroyd, and J. F. Gibbons: "Physical Electronics and Circuit Models of Transistors," John Wiley and Sons, Inc., New York, 1964.
3. Harbourt, Cyrus: Lumped Models for Negative-parameter Electrical Devices, *IEEE Trans. Circuit Theory,* vol. CT-12, no. 4, pp. 497–503, December, 1965.
4. Koehler, Dankwart: The Charge-control Concept in the Form of Equivalent Circuits, Representing a Link Between the Classic Large Signal Diode and Transistor Models, *Bell System Tech. J.,* pp. 523–575, March, 1967.
5. Lindholm, F. A. and D. J. Hamilton: A Systematic Modeling Theory For Solid State Devices, *Solid State Electronics,* vol. 7, pp. 771–783, Pergamon Press, New York, 1964.
6. Linvill, John G.: "Models of Transistors and Diodes," McGraw-Hill Book Company, New York, 1963.
7. Millman, J.: "Electronic Devices and Models," McGraw-Hill Book Company, New York, in preparation.
8. Sparkes, John J.: Device Modeling, *IEEE Trans. Electron Devices,* vol. ED-14, no. 5, pp. 229–232, May, 1967.

E-8 REFERENCES ON THE FORMULATION OF EQUATIONS OF MOTION OF NONLINEAR NETWORKS

1. Brayton, R. K. and J. K. Moser: A Theory of Nonlinear Networks, *Quart. Appl. Math.*, vol. 22, no. 1, pp. 1–33, July, 1964, and vol. 22, no. 2, pp. 81–104, July, 1964.

2. Chua, L. O. and R. A. Rohrer: On the Dynamic Equations of a Class of Nonlinear RLC Networks, *IEEE Trans. Circuit Theory,* vol. CT-12, no. 4, pp. 475–489, December, 1965.
3. Desoer, C. A. and J. Katzenelson: Nonlinear RLC Networks, *Bell System Tech. J.*, vol. 44, pp. 161–198, January, 1965.
4. Seshu, S. and M. B. Reed: "Linear Graphs and Electrical Networks." Addison-Wesley Publishing Company, Inc., Reading, Mass., 1961.
5. Stern, T. E.: On the Equations of Nonlinear Networks, *IEEE Trans. Circuit Theory,* vol. CT-13, no. 1, pp. 74–81, March, 1966.
6. Stern, Thomas E.: "Theory of Nonlinear Networks and Systems, An Introduction," Addison-Wesley Publishing Company, Inc., Reading, Mass., 1965.

E-9 REFERENCES ON NUMERICAL TECHNIQUES

1. Conte, S. D.: "Elementary Numerical Analysis," McGraw-Hill Book Company, New York, 1965.
2. Hildebrand, F. B.: "Introduction to Numerical Analysis," McGraw-Hill Book Company, New York, 1956.
3. Moursund, D. G. and C. S. Duris: "Elementary Theory and Application of Numerical Analysis," McGraw-Hill Book Company, 1967.
4. Ostrowski, A. M.: "Solution of Equations and Systems of Equations," 2d ed., Academic Press, Inc., New York, 1966.
5. Ralston, Anthony: "A First Course in Numerical Analysis," McGraw-Hill Book Company, New York, 1965.
6. Southworth, Raymond W. and Samuel L. DeLeeuw: "Digital Computation and Numerical Methods," McGraw-Hill Book Company, New York, 1965.
7. Todd, John: "Survey of Numerical Analysis," McGraw-Hill Book Company, New York, 1962.

E-10 REFERENCES ON COMPUTER-AIDED ANALYSIS

1. Calahan, Donald A: "Computer-aided Network Design," prelim. ed., McGraw-Hill Book Company, New York, 1968.
2. Huelsman, Lawrence P.: "Digital Computations in Basic Circuit Theory," McGraw-Hill Book Company, New York, 1968.

3. Jensen, Randall W. and Mark D. Lieberman: "IBM Electronic Circuit Analysis Program," Prentice-Hall, Inc., Englewood Cliffs, N.J., 1968.
4. Karplus, Walter J.: "On-line Computing," McGraw-Hill Book Company, 1967.
5. Korn, Granino A. and Theresa M. Korn: "Electronic Analog and Hybrid Computers," McGraw-Hill Book Company, New York, 1964.
6. Kuo, Franklin F. and James F. Kaiser: "System Analysis by Digital Computers," John Wiley & Sons, Inc., New York, 1966.
7. Kuo, Franklin F. and Waldo G. Magnuson, Jr.: "Computer Oriented Circuit Design," Prentice-Hall, Inc., Englewood Cliffs, N.J., 1969.
8. Levine, Leon: "Methods for Solving Engineering Problems," McGraw-Hill Book Company, New York, 1964.
9. Mischke, Charles R.: "An Introduction to Computer-aided Design," Prentice-Hall, Inc., Englewood Cliffs, N.J., 1968.

E-11 REFERENCES ON GENERAL NONLINEAR NETWORK PROPERTIES AND THEOREMS

1. Allanson, J. T.: On Nonlinear Constant-resistance Networks, *IEEE Trans. Circuit Theory,* vol. CT-13, no. 4, pp. 409–412, December, 1966.
2. Cherry, C.: Some General Theorems for Nonlinear Systems Possessing Reactance, *Phil. Mag.*, ser. 7, vol. 42, no. 333, pp. 1161–1177, October, 1951.
3. Cherry, C.: Classes of 4-Pole Network Having Nonlinear Transfer Characteristics but Linear Iterative Impedances, *Proc. IEE* (*London*), vol. 107B, no. 31, pp. 26–30, January, 1960.
4. Chua, L. O.: Equivalent Nonlinear Networks, *Proc. Midwest Symp. Circuit Theory, 8th,* pp. 20(0)–20(12), June, 1965.
5. Chua, L. O.: Graphical Analysis and Synthesis of Memoryless Nonlinear Networks, *Proc. IEEE Int. Conv. Record,* pt. 7, pp. 224–232, 1966.
6. Chua, L. O.: An Explicit Procedure for Determining the Δ-Y and Y-Δ Transformations for Nonlinear Networks, *Purdue*

Sixth Semiannual Research Summary, Purdue University, Lafayette, Ind., pp. 101–104, January–June, 1967.

7. Chua, Leon O.: Analysis and Synthesis of Multivalued Nonlinear Networks, *IEEE Trans. Circuit Theory,* vol. CT-15, June, 1967.
8. Chua, L. O.: Two New Theorems in Nonlinear Networks, *Midwest Symp. Circuit Theory, 11th,* University of Notre Dame, South Bend, Ind., pp. 340–351, May 13–14, 1968.
9. Desoer, C. A. and K. K. Wong: Constant Resistance One-ports Which Include Nonlinear Time-varying Elements, *IEEE Trans. Circuit Theory,* vol. CT-13, no. 4, pp. 403–412, December, 1966.
10. Donaldson, R. W.: Realization of Piecewise-linear Driving-point and Transfer Characteristics by RDS Networks, *IEEE Trans. Circuit Theory,* vol. CT-14, pp. 52–56, March, 1967.
11. Duinker, S.: Traditors, A New Class of Nonenergic Nonlinear Network Elements, *Philips Res. Rept.,* vol. 14, pp. 29–51, 1959.
12. Duinker, S.: Conjunctors, Another New Class of Nonenergic Nonlinear Network Elements, *Philips Res. Rept.,* vol. 17, pp. 1–19, 1962.
13. Kain, R. Y.: Diode Network Minimization, *IEEE Trans. Circuit Theory,* vol. CT-11, pp. 341–350, September, 1964.
14. Manley, J. M. and H. E. Rowe: General Energy Relations in Nonlinear Inductors and Capacitors, *Proc. Symp. Nonlinear Circuit Analysis,* Polytechnic Institute of Brooklyn, N.Y., vol. 6, pp. 399–407, April, 1956.
15. Millar, W.: Some General Theorems for Nonlinear Systems Possessing Resistance, *Phil. Mag.,* ser. 7, vol. 42, no. 333, pp. 1150–1160, October, 1951.
16. Millar, W.: The Nonlinear Resistive 3-Pole: Some General Concepts, *Proc. Symp. Nonlinear Circuit Analysis,* Polytechnic Institute of Brooklyn, N.Y., vol. 6, pp. 367–398, 1956.
17. Page, Chester H.: Frequency Conversion with Nonlinear Reactance, *J. Res. Nat. Bur. Std.,* vol. 58, no. 5, pp. 227–236, May, 1957.
18. Page, Chester H.: Harmonic Generation with Ideal Rectifiers, *Proc. IRE,* pp. 1738–1740, October, 1958.

19. Pantell, Richard H.: General Power Relationships for Positive and Negative Nonlinear Resistive Elements, *Proc. IRE*, pp. 1910–1913, December, 1958.
20. Penfield, Paul, Jr.: "Frequency-power Formulas," The M.I.T. Press, Cambridge, Mass., 1960.
21. Tellegen, B. D. H: A General Network Theorem, with Applications, *Philips Res. Rept.*, vol. 7, pp. 259–269, 1952.
22. Weber, Ernst: Nonlinear Physical Phenomena, *Proc. Symp. Nonlinear Circuit Analysis,* Polytechnic Institute of Brooklyn, N.Y., vol. 2, pp. 1–27, 1953.

E-12 REFERENCES ON APPROXIMATION TECHNIQUES

1. Akhiezer, N. I.: "Theory of Approximation," Frederick Ungar Publishing Co., New York, 1956.
2. Cheney, E. W.: "Introduction to Approximation Theory," McGraw-Hill Book Company, New York, 1966.
3. Davis, Philip J.: "Interpolation and Approximation," Blaisdell Publishing Company, Waltham, Mass., 1965.
4. Kantorovich, L. V. and V. I. Krylov: "Approximate Methods of Higher Analysis," Interscience Publishers (Division of John Wiley & Sons, Inc.), New York, 1964.
5. Krylov, V. I.: "Approximate Calculation of Integrals," The Macmillan Company, New York, 1962.

E-13 REFERENCES ON QUALITATIVE METHODS OF ANALYSIS

1. Cesari, Lamberto: "Asymptotic-behavior and Stability Problems in Ordinary Differential Equations," 2d ed., Academic Press, Inc., New York, 1963.
2. El'sgolc, L. E.: "Qualitative Methods in Mathematical Analysis," vol. 12, Translation of Mathematical Monographs, American Mathematical Society, Providence, R.I., 1964.
3. Krasnosel'skiy, M. A., A. I. Perov, A. I. Povolotskiy, and P. P. Zabreiko: "Plane Vector Fields," Academic Press, Inc., New York, 1966.
4. Lefschetz, Solomon: "Differential Equations–Geometric Theory," John Wiley & Sons, Inc., New York.

5. Nemytskii, V. V. and V. V. Stepanov: "Qualitative Theory of Differential Equations," Princeton University Press, Princeton, N.J., 1960.
6. Pliss, V. A.: "Nonlocal Problems of the Theory of Oscillations," Academic Press, Inc., New York, 1966.

E-14 REFERENCES ON ANALYTICAL TECHNIQUES

1. Bogoliuboff, N. N. and Y. A. Mitropolsky: "Asymptotic Methods in the Theory of Nonlinear Oscillation," English translation by Hindustan Publishing Corp., India, 1961.
2. Coddington, E. A. and N. Levinson: "Theory of Ordinary Differential Equations," McGraw-Hill Book Company, New York, 1955.
3. Cunningham, W. J.: "Introduction to Nonlinear Analysis," McGraw-Hill Book Company, New York, 1958.
4. Hale, Jack K.: "Oscillations in Nonlinear Systems," McGraw-Hill Book Company, New York, 1963.
5. Hayashi, Chihiro: "Nonlinear Oscillations in Physical Systems," McGraw-Hill Book Company, New York, 1964.
6. Krylov, N. and N. Bogoliubov: "Introduction to Nonlinear Mechanics," Annals of Mathematical Studies, no. 11, Princeton University Press, Princeton, N.J., 1947.
7. Minorsky, Nicholas: "Nonlinear Oscillations," D. Van Nostrand Company, Inc., Princeton, N.J., 1962.
8. Saaty, Thomas L. and Joseph Bram: "Nonlinear Mathematics," McGraw-Hill Book Company, New York, 1964.
9. Saaty, Thomas L.: "Modern Nonlinear Equations," McGraw-Hill Book Company, New York, 1967.
10. Stoker, J. J.: "Nonlinear Vibrations," Interscience Publishers, (Division of John Wiley & Sons, Inc.), New York, 1950.
11. Urabe, Minoru: "Nonlinear Autonomous Oscillations–Analytical Theory," Academic Press, Inc., New York, 1967.
12. Wasow, Wolfgang: "Asymptotic Expansions for Ordinary Differential Equations," Interscience Publishers (Division of John Wiley & Sons, Inc.), New York, 1965.

INDEXES

INDEX OF THEOREMS

Dynamic network solution theorems:
  capacitor-inductor-loaded isocline network, 836
  capacitor-loaded isocline network, 835
  complete-response, 783
  constant-slope capacitor-loaded network, 799
  constant-slope inductor-loaded network, 799–800
  explicit equilibrium state, 616
  exponential waveform solution, 701
  implicit equilibrium state, 615
  inductor-loaded isocline network, 836
  initial conditions, choice of, 588–589
  order of complexity, determination of, 590–591
  trajectory slope, 865
  zero-state response, 780

Equivalence theorems:
  current-source substitution, 331
  i-shift, 334
  Norton's, nonlinear, 328
  open-circuit substitution, 332
  short-circuit substitution, 332
  Thévenin's, nonlinear, 328
  v-shift, 332
  voltage-source substitution, 331
Existence theorems:
  Bendixon's, 875

Existence theorems:
  combined interpolation and approximation, 907
  DP plot, 245
  dual network, 208
  fundamental interpolation, 905
  monotonic solution, 646
  TC plot, 245
  uniqueness, 246
  Weierstrauss approximation, 906

Miscellaneous theorems:
  complementary network, 353
  Kuratowski's, 208
  positive-negative logic network conversion, 474

Realization and synthesis theorems:
  lattice network synthesis, 408
  scalor-reflector realization, 399
  scalor-rotator realization, 397

Symmetry theorems:
  flip-over symmetric network, 345
  rotationally symmetric isocline network, 878
  rotationally symmetric network, 346
  symmetric DP plot-TC plot, 350
  symmetric lattice network, 346
  vertically symmetric isocline network, 877

SUBJECT INDEX

Abrupt-junction diode, 32
AC-dynamic network, 226*n.*
AC-resistive network, 226*n.*
Active components, 387
Active resistor, 42, 63, 242
Addition, graphical, 920
Algebraic equation, 155*n.*
Algorithm, 179
Almost periodic waveform, 39
Ampere, 4, 53
Amplification factor, 508
Amplifier:
  complementary symmetric, 356–357
  dc, 355–356
  differential input, 149
  logarithmic, 238–241
  operational, 109–110
    applications of, 115–128, 145
    in TC plot synthesis, 405–407
  parametric, 34, 894
  pulse, 733
  push-pull, 357
Amplitude modulation (*see* Modulation, amplitude)
Analog circuit, 117–128
Analog computer, 126–127
Analog-to-digital conversion, 431–433
Analysis:
  of first-order autonomous networks, 636–696
  of first-order nonautonomous networks, 778–822
  of first-order switching networks, 697–724
  graphical methods of, 253–324
  of *n*th-order autonomous networks, 880–897
  piecewise-linear methods of, 526–580
  of second-order autonomous networks, 823–879
AND gate, 111–112, 148, 471–478
Aperiodic waveform, 426
Approximate analytic approach, 895
Approximation, 905
Arc length, 910
Assumed reference convention, 7–9
Assumed states, method of, 577
Astable multivibrator, 726–730
Asymptotic stability, 624
Automatic comparison circuit, 282–285
Automatic sorting circuit, 280–282
Autonomous network, analysis of, 195
    first-order, 636–696
    *n*th-order: approximate analytic approach, 895
      heuristic approach, 885–890
      numerical-integration approach, 881–883
      piecewise-linear approach, 884–885
      qualitative approach, 890–895
    second-order: analytic methods, 843-847
      isocline network method, 833–842
      phase-plane technique, 823–833
      piecewise-linear technique, 847–859
Avalanche diode, 20
Avalanche voltage, 276

Back diode, 21
Back-to-back parallel connection, 271
Back-to-back series connection, 350
Back-to-front parallel connection, 350–351
Back-to-front series connection, 267
Balancing property of bridge circuit, 284*n.*
Barium titanate dielectric, 70
Base-and-peak clipper, 428
Base clipper, 427
Basic network configurations, 255–266
Beat frequency, 27–28
Bendixon's theorem, 875–876
Biasing:
  three-terminal resistor, 374–382, 413
    sensitivity considerations, 382–386
  two-terminal resistor, 369–373, 413
Bidirectional switch, 452
  (*See also* Switch)
Bilateral resistor, 23, 343
Binary logic, 738
Bistable multivibrator, 737–746
Black box, 11–17, 423
Bootstrap circuit, 776
Branch, 156, 158
Breakdown diode, 20
Breakpoint, 566, 905
Bridge circuit, 285
Brute force method, 794
Building blocks, 387

Canonic form, 526–530, 573
Capacitance, 31
  incremental, 32
  self- and mutual, 104*n.*
Capacitor:
  controlled, two-terminal, 67, 78
  dual of, 204
  linear, 30–31, 104
  nonlinear, 31–32
    applications of, 32–35
  power, average power, energy in, 42–45
  three-terminal, 103–104
  time-varying, 49–50
  two-terminal, 16, 30–35
  unicursal, 913
Capacitor-inductor-loaded isocline network theorem, 836
Capacitor-loaded isocline network theorem, 835
Carrier, 49, 60, 814
Cascade connection of three-terminal resistors, 313–315
Catching diode, 774
Cathode follower, 431*n.*
Cell, 792
Characteristics:
  of three-terminal resistor, 86–98
  of two-terminal black box, 11–17
Characterization, 489
Charge, 3, 5, 53
  conservation of, 654, 656
Chopper, 461
  synthesis of, 461–463
C-L mutator, 142–143
Clamping circuits, 809
Classification:
  of curves, 19–23
  of dynamic networks, 583–585, 626
  of networks, 152–153, 213
Clipper, 299, 427–431, 734
Closed trajectory, 609
Coincidence gate, 466–467
Combined interpolation and approximation theorem, 907
Commensurate frequencies, 27, 890, 892
Common-emitter amplifier, 415
Comparison circuit, 282–285
Compensation, 72, 433–437, 443
Complementary network, 353–355
Complementary-network theorem, 353–354, 361, 379, 389
Complementary symmetric amplifier, 356
Complete response, 781–783
Complete-response theorem, 783
Complexity of dynamic network (*see* Order of complexity)
Composite characteristics of three-terminal resistors, 311–315
Composition, 75–78, 292, 410, 565
  graphical, 921–923
Computer programs for analysis of resistive networks, 558–562, 573–574

Concave resistor (*see* Resistor, concave and convex)
Conductance, 17
  ac, 57
  dc, 57
  incremental, 641
  short-circuit, 94
Conductance representation, 89, 94, 144
Conductance-resistance representation, 92, 94, 144
Connected graph, 158
Connected network, 157–158
Conservation:
  of charge, 654, 656
  of flux linkages, 656
Constant-current diode, 20, 296, 387, 395, 418
Constant-slope capacitor-loaded network theorem, 799
Constant-slope curve, 794
Constant-slope inductor-loaded network theorem, 799–800
Constant-slope method, 793–806, 819
Constant-slope network, 797–800
Constraints on multiterminal elements, 108–115
Contradiction method, 474
Controlled elements, 62–83
  applications of, 68–70
  two-terminal, 62–67, 78
Controlled sources, 70–72, 118–121, 401, 495–502, 526, 537, 932–937
Controlling variable, 62
Convergence, rate of, 185
Convex resistor (*see* Resistor, concave and convex)
Convolution integral, 780*n.*
Cotree, 158, 214
Coulomb, 3, 53
Coupling network, 117
Cowan modulator, 463
Cross-coupled multivibrator (*see* Eccles-Jordan circuit)
Cryotron, 63, 65
Current, 4, 53
Current direction reference, 7–9, 17*n.*, 53
Current divider, nonlinear, 290
Current-inversion negative-impedance converter (INIC), 137
Current-limiting diode, 18*n.*
Current ratio-voltage ratio representation, 93, 94, 144
Current-source substitution theorem, 331, 335
Curve generator, 495
Curve tracing circuit, 929–931
Cut set, 163, 214
  fundamental, 165, 214

Damping constant, 867
Darlington circuit, 340–342
Dc dynamic network, 226*n.*
Dc resistive network, 226*n.*
Decomposition method, 393
Delay line, 734
Delay time, 6, 77
Demodulator, 814
Describing function, 58, 59
Design, 368
Detector, 814
Differential input amplifier, 149
Differentiation, graphical, 924–926
Differentiator, 121–122, 784–785
Diode, 20–22
  abrupt-junction, 32
  catching, 774
  field-effect, 18*n.*
  ideal, 23–24
  junction, 18
  light-emitting, 75
  tunnel, 19
  vacuum, 18
  varactor, 35
  zener, 20, 296, 387, 395, 418, 728
Director, 793
Discrete signal, 431
Distortion:
  linear, 805
  nonlinear, 804–805
Distributed network, 6*n.*
Divider, nonlinear: current, 290
    in TC plot synthesis, 402–405, 562–564
    voltage, 286–289

Division, graphical, 920–921
Domain, 899
Double load-line method, 263–264
  (*See also* Template-double load-line method)
Double-shot multivibrator (*see* Bistable multivibrator)
DP plot, 228–232, 246
  applications of, 273–285, 306–311
  and equivalence, 327–330
  graphical determination of: for networks with three-terminal resistors, 301–311
    for series-parallel networks, 266–273
  horizontal and vertical segments in, 276–277
  piecewise-linear determination of, 539–553
  and symmetry, 348–353
  synthesis of, 386–401, 413
DP plot theorem, 245, 247
Driving-point plot (*see* DP plot)
Driving-point terminals, 231, 232
Dual elements, 203–204
  self-, 221
Dual network, 202–213
  algorithm for, 209
  reference direction for, 206
  self-, 221
Dual network existence theorem, 208
Duality, 201–213
Dynamic characteristics, 15
Dynamic network, 153, 213, 583
  analysis of: first-order autonomous, 636–696
    first-order nonautonomous, 778–822
    first-order switching, 697–724
    *n*th-order autonomous, 880–897
    second-order autonomous, 823–879
  classification of, 583–585
  equations of motion for, 188–197
  order of complexity of, 585–593
  principles of analysis, 593–595
  synthesis of: first-order multivibrators, 725–752
    first-order time-base generators, 753-777
Dynamic route, 671

Ebers-Moll equations, 489–490
Eccles-Jordan circuit, 579, 878
Edge of element (branch), 156
Eigenvector, 863
Elements:
  laws of, 153–155, 213
  multiterminal, 84–151
    characterization of, 84–85
    constraints on, 108-115
    three-terminal elements: capacitors, 103–104
      inductors, 104–105
      resistors, 86–103
  nonlinear parametric representation of, 913–918
Elimination, graphical, 287, 411, 923–924
Emitter follower, 431*n*.
Empty set, 546*n*.
Energy, 3, 5, 38–47, 53
Energy dissipation, 53
Energy-storage element, 45, 47, 53
Energy-storage subnetwork, 593
Envelope detector, 814
Equalization (*see* Compensation)
Equations of motion, 152–221
Equilibrium state, 613–622, 627
  for first-order networks, 637–640
  stability of: first-order network, 622–626
    second-order network, 859–863
  and trajectory: first-order network, 620–622
    second-order network, 859–873
  virtual, 661, 857
  (*See also* Singular point)
Equivalence, 325–367
Equivalent input signal, 452
Equivalent load specification, 565
Equivalent networks, 325–327
Equivalent source property, 328
Euler algorithm:
  for *n* differential equations, 199–201, 215, 883
  for one differential equation, 196–199
Exclusive OR gate, 148
Explicit equilibrium state theorem, 616, 627
Exponential waveform solution theorem, 701, 718
Extraneous solution, 331

Farad, 31
Feedback network, 117
FET (*see* Field-effect transistor)
Field-effect transistor, 102, 308
  models of, 517
  time-base generator circuit, 770–771
Filter, 26
Final value, 662*n.*
Fixed-point algorithm, 181, 215
Fixed-point concept, 180–184
Flip-flop (*see* Bistable multivibrator)
Flip-over symmetric-network theorem, 345
Flip-over symmetry, 343, 361
Flux, magnetic, 4
Flux linkage, 4, 5, 53
  conservation of, 656
Flyback time, 755
Focus:
  stable, 867
  unstable, 868
Four-layer diode, 22
Free-running multivibrator (*see* A stable multivibrator)
Free-running time-base generator (*see* Time-base generator, free-running)
Frequency:
  natural, 867
Frequency conversion, 37
Frequency division, 33, 768
Frequency doubler, 444
Frequency mixing, 27–28
Frequency multiplication, 25–27
Full-wave rectifier, 337
Function, 901
  and inverse, 901–904
  mathematical representations of, 904–913
  monotonic, 903
  multivalued, 902
  one-to-one, 902
  piecewise-linear, 905
Functional equation, 155*n.*
Functional network, 422, 433
  synthesis of: ac-resistive, 449–488
    dc-resistive, 422-448
Fundamental cut set, 165, 214
Fundamental interpolation theorem, 905
Fundamental loop, 160, 214

Gate:
  AND, 111–112, 471–474
  coincidence, 466–467
  exclusive OR, 148
  NAND, NOR, and NOT, 471–472, 476–479
  OR, 111–113, 471–472, 474–476
  transmission, 468–469
Gated oscillator, 736*n.*
Generalized Kirchhoff current law, 164–165
Global model (*see* Model, global)
Global stability, 623
Glow tube, 22
Graph of network, 156
Graphical analysis, 253–324
  of non-series-parallel networks, 334–342
Graphical elimination method, 287–288, 289, 411
Graphical synthesis (*see* Synthesis)
Graphical techniques for basic mathematical operations, 919–928
  addition and subtraction, 920
  composition, 921–923
  differentiation, 924–926
  elimination, 923–924
  integration, 926–928
  multiplication and division, 920–921
Ground, 85
Gyration conductance, 136
Gyrator, 135–136
  circuits for simulating, 940–942

Half-wave rectifier, 294
Harmonic, 37
Henry, 35
Heun algorithm, 219
Heuristic approach, 885–886
Holding current, 457
Hooke's law, 154
Horizontal flip-over symmetry, 356
Horizontal segments in DP plot, 276–277
Horizontal symmetry, 351
Hybrid form, 605
Hybrid parameters, 95
Hysteresis, 36, 917

i-shift theorem, 334, 336, 341, 502

Ideal diode, 23–24
Image, 903
Implicit equilibrium state theorem, 615, 627
Impulse, 818*n.*
Impulse response, 780*n.*
Incomplete model, 650–656
Independence:
  of equations, 160–166
  of variables, 16, 84–85, 159
Independent initial condition, 585
Independent KCL equation criteria, 166, 214
Independent KVL equation criteria, 160, 214
Independent KVL mesh equation criteria, 160
Inductance, 35, 36
  incremental, 37
  self- and mutual, 105*n.*
Induction, mathematical, 158*n.*
Inductor:
  controlled two-terminal, 67, 68
  dual of, 204
  linear, 35–36, 105
  nonlinear, 36–37
    applications of, 37–38
  power, average power, energy in, 45–47
  three-terminal, 104–105
  time-varying, 50–51
  two-terminal, 16, 35–38
  unicursal, 913
Inductor-loaded isocline network theorem, 836
Inertia postulate, 651, 732
Initial condition, 197, 586–589, 627
Input-characteristic curves, 86
Input load line, 262
Instability problem, 388
Integrated circuits, 7, 31, 110, 236, 356, 388
Integration, graphical, 926–928
Integrator, 122–123, 783–784, 931
Interconnection, laws of, 153–157, 213
Internal constraints, 108–113
Interpolation, 905
Intersection, 546
Invariant network, 342
Inverse function, 901
Inverse machine, 899
Inverse relation, 900
Inverter, phase, 120
Isocline, 823–828
Isocline network, 833–836
Isocline network method, 833–842
Isolation block for integrated circuits, 236–238
Isolation box, 119
Isolation transformer, 114
Iterative linear network, 528, 573
Iterative Norton equivalent network, 528, 573
Iterative piecewise-linear method (*see* Piecewise-linear method)
Iterative Thévenin equivalent network, 528, 573

Jump phenomenon, 652–656
Junction diode, 20

Kirchhoff current law, 154
  in generalized form, 164, 214
  and independent equations, 163–166
Kirchhoff voltage law, 154
  and independent equations, 160-162
Kuratowski's theorem, 208

Ladder network, 290, 392–395
Latching switch, 456–458
Lattice network, 337
  symmetric, properties of, 351–352
Lattice network synthesis theorem, 408, 565
Laws:
  of elements, 153–154
  of interconnection, 153–154
Liapunov, 624
Lienard's method, 874–875
Light-emitting diode, 75
Limit cycle, 610, 875
Limiter, voltage, 296–297
Limiting, 28–30
Linear distortion, 805
Linear field, 824*n.*
Linear network, 152–153
Linear transformation converter (LTC), 151
Link, 158
Load, 234

- Load line, 257
- Load-line method, 256–257, 316
  - double, 263–264, 316
- Loading effect, 234
- Local model, 494
- Local stability, 623
- Locus of operating points, 228
- Logarithmic amplifier, 238–241
- Logic, 469
  - positive and negative, 473–474
- Logic elements, 110–113
  - synthesis of, 469–479
- Lossless element, 42–47, 868
- Lumped network, 6

- Manley-Rowe equations, 892–893
- Mapping, 903
- Mathematical induction, 158*n.*
- Matrix transmission gate, 468–469
- Maximum power hyperbola, 372
- Meca program, 559–562, 574
- Memory (*see* Storage)
- Memoryless network, 153*n.*
- Mesh, 161–162
- Metal oxide semiconductor, 31
- Mho, 17
- Mode, natural or normal, 844
- Model, 51-52, 489–491, 520
  - dynamic, 51–52, 618
  - global, 494
    - for controlled resistors, 494–497
    - of practical three-terminal resistors, 508–520
    - for three-terminal resistors, 497–503
  - incomplete, 650–656
  - incremental linear, 494*n.*
  - local, 494
  - modification of: half-plane constraints, 503–504
    - quadrant constraints, 505–507
  - and parasitic elements, 656–659
  - realistic and incomplete, 649–652
- Modulation, amplitude, 48–49, 59
  - detection of envelope, 814–817
  - synthesis of modulator, 461–465
- Modulation index, 814
- Monostable multivibrator, 725, 730–737, 746
- Monotonic charging property, 807–811
- Monotonic curve, 388–389
  - DP plot, synthesis of, 388–396
  - (*See also* Nonmonotonic curve)
- Monotonic function, 903
- Monotonic solution theorem, 646
- Multiplication, graphical, 920–921
- Multivalued curve, 23, 270, 730, 737, 840, 902
- Multivibrator, 513
  - first-order, 725, 746
    - astable, 726–730
    - bistable, 737–746
    - monostable, 730–733
      - applications of, 733–734
      - conditions on triggering pulse amplitude, 735–736
      - conditions on triggering pulse width, 736–737
- Mutator, 31, 138, 146
  - circuits for simulating, 940–942
  - kinds of: *C-L*, 142–144
    - *R-C*, 140–142
    - *R-L*, 138–140
- Mutual capacitance, 104*n.*
- Mutual inductance, 105*n.*

- *N*-channel FET, 102
  - model of, 517
- *N-p-n* transistor, 99
  - model of, 512–515
- NAND gate, 471–472, 476–479
- Negative-impedance converter, 138
  - circuits for simulating, 937–938
  - current-inversion, 137–138
  - voltage-inversion, 136–137
- Negative logic, 473–474
- Negative resistance, 19, 726
- Neon bulb time-base generator, 760–762
- Network, 152–153
  - autonomous, 195
  - basic configurations of, 255, 260
  - connected, 157–158
  - dual, 202–203, 215

Network:
  dynamic, 153, 213, 583
  equivalence, 325–327, 361
  functional, 422, 433
  graph, 156
  invariance, 342
  model, 489–491
  planar, 161, 215
  resistive, 153, 213
  series-parallel, 290–293
  symmetry, 342–353
  topology, 155, 157–166, 213–214
Network analysis (*see* Analysis)
Network synthesis (*see* Synthesis)
Newton-Raphson method, 185–187, 215
Node, stable and unstable, 865–866
Noise, 429*n.*
Nonautonomous network, 195
  first-order, 778
    applications of, 807–819
    constant-slope method, 793–806
    linear network analysis, 779–786
    piecewise-linear analysis, 786–792
Nonlinear distortion, 804–805
Nonlinear network, 152–153
Nonlinear Norton's theorem, 328–330
Nonlinear oscillation, 622*n.*
Nonlinear Thévenin's theorem, 328–330
Nonmonotonic curve, synthesis of, 396–401
Nonplanar network, 161
NOR gate, 471–472, 476–479
Normal-form equations, 191–196
  first-order networks, 595–599
  *n*th-order networks, 604–606
  second-order networks, 599–604
Normalized isocline network, 876
Norton equivalent network, interative, 528, 573
Norton's theorem, nonlinear, 228
NOT gate, 471–472, 476–479
Null detection circuit, 282–285
Numerical-integration approach, 881–883
Numerical solution:
  of functional equations, 179–188, 215
  of nonlinear differential equations, 196–201, 215, 219, 220
Nyquist theorem, 482

Offset voltage, 127
Ohm, 17
Open circuit, 17, 23
Open-circuit resistances, 94
Open-circuit substitution theorem, 332, 343, 346
Open-circuit voltage, 329
Open trajectory, 609
Operating point, 225–228, 246
  applications of, 273–285
  and equivalence, 330–332
  graphical determination of, 254–266
  piecewise-linear determination of, 530–538
Operating-point paradox, 243–244
Operating-point problem, 227–228
Operational amplifier (*see* Amplifier, operational)
Optical coupling, 76
OR gate, 111–113, 471–472, 474–476
Order of complexity, 585–593, 626
  determination of, 589–592
Oscillatory solution, 652–656
Output characteristic curves, 87
Output load line, 262
Overdamped network, 867
Overload protection, 297

P-channel FET, 102
  model of, 517
P-n-p transistor, 99
  model of, 517
Paradox, 29, 243
Parallel combination technique, 267, 271–273
Parallel commutative property, 327
Parallel decomposition technique, 393
Parametric amplifier, 34, 894
Parametric approach, 909
Parametric representation of nonlinear elements, 913–918
Parasitic elements, 52, 656–659, 703, 744, 888
Passive components, 387
Peak clipper, 427
Peak current, 279*n.*
Pentode (*see* Vacuum pentode)
Per-unit ripple, 813

Periodic signal, 39, 425–426
Phase inverter, 120
Phase plane, 609
Phase-plane technique, 823–833
Phase portraits, 823–828, 871–873
Photocell, 65
Photodiode, 63
  equation of v-i curves of, 66
Phototube, 63
π-network model, 501–503
Piecewise-linear function, 905
Piecewise-linear method, 526–530
  in analysis: computer programs, 558–562
    DP plots, 539–553
    first-order autonomous networks, 659–689
    first-order nonautonomous networks, 786–792
    nth-order autonomous networks, 884–885
    operating-points, 530–538
    second-order autonomous networks, 844–859
    TC plots, 553–558
  in synthesis, 562–573
Pivoting point, 926
Planar network, 161
Pliers-type entry, 232
Port contraints, 113–115
Positive logic, 473–474
Positive-negative logic conversion theorem, 474
Potential difference (*see* Voltage)
Power, 5, 38–47, 53
  significance of sign, 9, 53
Power gain, 894
Power measuring device, 79
Power rating, maximum, 241–243, 371–372
Power-transfer plot, 241–243, 246
Pulse amplifier, 733
Pulse compressor, 297–298
Pulse generator, 298–301
Pulse-sharpening property, 817–818
Pulse-smoothing property, 811–817
Pulse stretcher, 812
Pulse synchronization method, 764
Pump, 894
Push-pull amplifier, 357

Qualitative methods, 607, 636, 842, 890–895
Quasi-resistive network, 424, 778

Range, 899
Rectification, 24–25
Rectifier, 24–25
  filtered, 810
  full-wave, 337
  half-wave, 294–296
Redundant element property, 327
Reference convention, 7–9, 206
Reflector, 128–129, 134–138, 145, 399
  circuits for simulating, 939–940
Regular point, 827
Regulation, 437–443
Relation, 899
  inverse, 900
Relaxation oscillator, 306–308, 727
Representations of a three-terminal resistor:
  forms of, 89–95
  graphical transformation of, 95–96
  transformation of, to another common terminal, 97–98
Resistance, 17
  ac, 56
  dc, 56
  incremental, 642
  negative, 19
  open-circuit, 94
Resistance-conductance representation, 92, 94, 144
Resistance representation, 92, 94, 144
Resistive network, 153, 213
Resistive subnetwork, 593
Resistor:
  active, 42, 53
  bilateral, 23, 343
  complementary, 353
  concave and convex, 273–276
  controlled: models of, 494–497
    two-terminal, 62–66, 78
  dual of, 203–204
  linear, 17-18
    negative, 17
    positive, 17

Resistor:
  multiterminal used as three-terminal controlled, 107–108
  nonlinear, 18–19
    applications of, 24–30
  passive, 42, 53
  power, average power, energy in, 37–42
  temperature-dependent, 63–66, 436
  three-terminal, 86–89
    biasing, 374–382
    cascade connection of, 313–315
    composite characteristics of, 311–315
    criteria for thermal runaway, 384–386
    models of, 497–503, 508–520
    practical examples of, 98–103
    representations of (*see* Representations of a three-terminal resistor)
    sensitivity considerations, 382–386
  time-varying, 47–49
  tunnel, 21
  two-terminal, 16–30
    biasing, 369–373
    criterion for thermal runaway, 373
  unicursal, 913
Restoration time, 755*n.*
Retrace time, 755
Return time, 755*n.*
Ring modulator, 465
Ripple, 813
Rise time, 704
Rotational symmetric isocline network theorem, 878
Rotational symmetric-network theorem, 346
Rotational symmetry, 343, 361
Rotator, 128–129, 131–134, 145, 397, 496, 728
  circuits for simulating, 939
Runge-Kutta algorithm, 220

Saddle point, 886
Sampling circuit, 461
Saturation, 110*n.*, 115
Saturation current in diode, 237, 238
Saturation region, 297
Sawtooth waveform, 308
Scaling circuit, 118–121
Scalor, 128–131, 145, 308–311
  circuits for simulating, 938–939
Scalor-reflector realization theorem, 399
Scalor-rotator realization theorem, 397
Schmitt trigger, 559–562, 722
SCR (*see* Silicon-controlled rectifier)
Segment-by-segment method, 389
Selenium diode, 20
Self-capacitance, 104*n.*
Self-dual, 221
Self-inductance, 105*n.*
Self-triggered time-base generator, 771–774
Semiautomatic general analysis program, 558–559, 573–574
Sensitivity, 371
  switching, 454, 479
  in three-terminal biasing circuits, 382–386
Separatrix, 796
Sensing element (*see* Transducer)
Series-combination technique, 267–271
Series commutative property, 327
Series configuration of switch, 450
Series-decomposition technique, 393
Series modulator, 463
Series-parallel network, 253, 290–293
Set, 898
Set-processing machine, 898
Shifting theorems, 332–334
Short circuit, 17, 23
  principle of virtual, 118
Short-circuit conductances, 94
Short-circuit current, 329
Short-circuit substitution theorem, 332
Shunt configuration of switch, 450
Shunt modulator, 463
Sideband frequency, 27*n.*
Sidebands, 27*n.*
Signal, 894
Signal-processing circuits, 294, 423, 807–819
Silicon-controlled rectifier, 102
  in high-sensitivity latching switch, 456–458
Simulation method, 125–127, 196
Simultaneity postulate, 6, 53
Single-shot multivibrator (*see* Monostable multivibrator)

Singular point, 614*n.*, 827, 859–873
  (*See also* Equilibrium state)
Skeleton network, 424
Slope of approach, 861
Small variation postulate, 25
Solar cell, 63
  equation of *v-i* curves of, 66
Soldering-iron entry, 232
Solion liquid diode, 21
Solution of the network, 154
Sorting circuit, 280–282
Source, 9
  charge, independent, 11
  controlled, 70–78
    circuits for simulating, 932–937
    controlled by electrical variables, 72–75
    using operational amplifier, 118–121
  current, 9–11
    independent, 11
  dc, 23
  flux-linkage, independent, 11
  voltage, 9–11
    independent, 10–11
Spiral case, 847, 860, 867
Square-law device, 238–241
Square-wave generator, 115–116, 308, 728–729
Stability of equilibrium states, 622–627
  asymptotic, 624
  criteria for first-order network, 637–643, 689
  definition of: for *n*th-order network, 626
    for second-order network, 624–625
  global (stability-in-the-large), 623
  local (stability-in-the-small), 623
Stable focus, 867
Stable spiral, 867
State space, 607
  trajectory of solutions in, 605–613
State variables, 191–196, 586
Static characteristics, 15
  of a two-terminal black box, 15–17
Steady-state behavior, 607, 782*n.*
Step size, 197
Storage, 37–38, 738
Strictly monotonically increasing curves, 19, 22
Subharmonic, 33, 37, 444
Subnetwork, 152–153
  energy-storage, 593
  resistive, 593
Substrate, 236
Subtraction, graphical, 920
Summing circuit, 123–125
Superconducting tunnel junction, 22–23
Supernode, 84, 165
Superposition, principle of, 819
Sweeptime, 755
Switch:
  electronically controlled, 449, 479
    characteristics of, 449–452
    synthesis of: multicontrolled, 465–469
      single-controlled bidirectional, 458–461
      single-controlled unidirectional, 452–458
  series, 449, 479
  shunt, 449, 479
Switching line, 788, 853
Switching network, first-order, 697–698, 718
    analysis of: linear, 698–707
      nonlinear, 708–710
      with one nonlinear inductor or capacitor, 711–718
Switching plane, 789
Switching sensitivity, 454, 479
Switching signal, 449
Switching space, 792
Switching transistor, 455*n.*
Symmetric DP-TC plot theorem, 350
Symmetric lattice network properties, 351
Symmetric lattice network theorem, 346, 408–411, 569
Symmetry, 342–353, 361
  applications of: determination of TC and DP plots, 348–353
    operating-point problem, 343–348
Synchronization of free-running time-base generator, 762–763
Synchronous phase detection, 49, 61
Synthesis:
  of functional networks: ac-resistive, 449–488
    dc-resistive, 422–448
  graphical methods of, 368–369
    of DP plots, 386–401
    of jointly prescribed curves, 408–412

Synthesis:
  graphical methods of: of operating points, 369–386
    of TC plots, 401–408
  of models: controlled resistors, 494–497
    three-terminal resistors, 497–507
  of multivibrators, first-order, 725
      astable, 726–730, 746
      bistable, 737–746
      monostable, 730–737, 746
  piecewise-linear methods of: for jointly prescribed curves, 565–573
    for TC plots, 562–573
  of time-base generators, first-order, 753–759
    free-running, 759–768
    triggered, 768–774

T-network model, 498–502
Taylor algorithm, 219
Taylor series expansion, 871
TC plot, 228–235, 246
  applications of, 294–301, 306–311
  graphical determination of: for networks with three-terminal resistors, 301–311
    for series-parallel networks, 286–294
  piecewise-linear determination of, 553–558
  and symmetry, 343–348
  synthesis of, 401–408, 413
    by piecewise-linear method, 562–573
TC plot theorem, 245, 247
Temperature coefficient, 65, 373
  (*See also* Sensitivity)
Template-double load-line method, 303–306, 316
Template method, 286, 289
Terminal state, 653
Tetrode, 107–108
Thermal coupling, 77
Thermal runaway, 372, 413
  criteria for three-terminal resistors, 384–385
  criterion for two-terminal resistors, 373
Thermistor, 63–66
  resistance of, as function of temperature, 65–66
Thévenin equivalent network, 528, 573
Thévinin's theorem, nonlinear, 228
Threshold detector, 116–117
Threshold value, 277–279
Thyratron, 99
Tilt angle, 433
Time-base generator, first-order, 753–759, 774
    free-running, 755–775
      neon bulb circuit, 760–762
      synchronization of, 763–768
      synthesis of, 759–763
      unijunction transistor circuit, 762–763
    self-triggered circuit, 771–774
    synthesis of, 768–774
    triggered, 757, 774
      FET circuit, 770–771
Time constant, 661, 675
Time-invariant elements, 47
Time scaling:
  first-order networks, 644–649
  second-order networks, 828–833
Time-translation property, 844*n.*
Time-varying elements, 47–51, 62*n.*
Tolerance, 765
Topology, 155, 157–166, 213–214
Toroidal coil, 67, 105
Trace, 869
Trajectory, 608
  closed, 609
  and equilibrium states: first-order network, 620–622
    second-order network, 859–873
  of solutions in state space, 606–613
  time scaling of: first-order networks, 644–649
    second-order networks, 828–833
Trajectory slope theorem, 865
Transcendental equation, 155*n.*
Transducer, 71–72, 78, 436
Transfer-characteristic plot (*see* TC plot)
Transfer terminals, 233
Transformer, 114–115
  isolation, 114–115
  three-port ideal, 114, 459
  three-winding, 359
  two-port ideal, 114
Transient response, 782*n.*

Transistor, 99
  models of: *n-p-n,* 512–516
    p-n-p, 517
Transistor switch, 454–456
Transition state, 653
Transitive relation property, 327
Transmission gate, 468–469
Transmission line, 5
Transmission parameters, 95
Tree, 158, 213
Trigger diode, 22
Triggered time-base generator (*see* Time-base generator, first-order, triggered)
Triggering source, 725, 730
Triggering techniques, 730–745, 768–774
Triode (*see* Vacuum triode)
Truncation error, 198
Truth table, 148, 471
Tuning, electronic, 34–35
Tunnel diode, 19, 21
Turns ratio, 114

Underdamped network, 868
Unicursal curve, 909
Unicursal element, 913
Unidirectional switch, 452
  (*See also* Switch)
Unijunction transistor, 102
  time-base generator circuit, 762–763
Uniqueness hypothesis, 343–344
Uniqueness theorem, 246, 247, 343*n*., 408, 537
Up-converter, 894*n*.

V-shift theorem, 332–334, 341, 351, 376, 460, 478, 499, 763
Vacuum diode, 18, 20
Vacuum pentode, 108
  models of, 510–511
Vacuum triode, 99
  models of, 508–510
Valley current, 279*n*.
Van der Pol equation, 824
Varactor diode, 35, 435
Varistor, 18, 21
Vector field, 794
Vertical-position transducer, 433
Vertical segments in DP plot, 277
Vertically symmetric isocline network theorem, 877
Virtual equilibrium state, 661
Virtual short circuit, principle of, 118, 406
Volt, 4, 53
Voltage, 3, 5, 53
  reference polarity, 7–9, 17*n*., 53
Voltage comparator (*see* Threshold detector)
Voltage-controlled resistor, 19
Voltage divider, nonlinear, 286–289
Voltage doubler, 810–811
Voltage gain, 110*n*.
Voltage-inversion negative-impedance converter (VNIC), 136–137
Voltage limiter, 296
Voltage ratio-current ratio representation, 92–94, 144
Voltage regulation, 297, 324
Voltage-source substitution theorem, 331
Voltage-stabilizer network, 167
Vortex, 867

Waveforms:
  generation of, 424–425, 778
    aperiodic, 426–427
    periodic, 425–426
Waveshaping, 427–433, 443, 778
Weber, 4, 53
Weierstrass approximation theorem, 906
Work (*see under* Energy)

Zener breakdown, 276
Zener diode, 20, 296, 387, 395, 418, 728
Zero-input response, 782
Zero-state response, 779–781
Zero-state response theorem, 780

This book was set in Times Roman by Graphic Services, Inc., and printed on permanent paper and bound by The Maple Press Company. The designer was Elliot Epstein; the drawings were done by Joseph Buchner. The editors were Michael Elia and Peter R. Karsten. Peter D. Guilmette supervised the production.

10804